On vocal assemblages

On Vocal Assemblages: From Edison to Miku, *Contemporary Music Review*, 36

Nick Prior, University of Edinburgh

[2nd and final draft: received 4.9.17, 9,300 words]

Introduction

Two scenes. In the final act of the 1947 film adaptation of Graham Greene's novel *Brighton Rock*, a message recorded on vinyl by the deceased gangster Pinkie, courtesy of a "record your voice" entertainment booth, is played back on a gramophone to his widow, Rose. In an act of malfunction that she fails to register, the record loops back on itself, declaring the words she is yearning to hear: "What you want me to say is: I love you….I love you…I love you…I love you…". Had the record not skipped, she would have discovered a very different message: "…but here's the truth, I hate you, you little slut, you make me sick". No matter. She is reassured by the vocal declaration, and as the camera zooms in on a crucifix, the illusions of love and redemption are upheld.

Fast forward to March 2014 and to the murder trial of the South African athlete, Oscar Pistorius. The athlete is accused of shooting dead his girlfriend, Reeva Steenkamp, on Valentine's Day 2013 and in a dramatic moment in court, one of Pistorius' neighbors testifies that she heard two distinct voices arguing on the night of the murder, one male and one female. At the moment of impact, the latter takes the form of a terrified scream, the neighbour says. In defence, Pistorius' lawyers state that the two voices are, in fact, one and the same. They are the shouts for help of the accused who, they say, "screams like a girl". And in any case, they continue, the existence of adjacent buildings and other intervening noises, such as those of wild animals, would have made it impossible to correctly identify the source of any voice.

In both examples, the voice carries with it the promise of and desire for truth: as witness and vehicle of love and murder, of identity and the body. Yet it turns out that this truth has unstable and contingent moorings. It is subject to the whims of malfunctioning technologies and inter-subjective misinterpretations that are made all the more fuzzy with spatio-temporal distance and intervening conditions such as buildings, non-human noises and gendered ideologies. What is at stake is the status of a sonic object imbued with special qualities but which simultaneously serves as an index of modern uncertainties about where the voice really resides: in the person, in the machine, in the murky spaces between addressor and addressee?

Despite being seen as the *sine qua non* of communicative primacy, voices are neither fixed nor stable. They can move us but they also move, comfort but also haunt us, entertain but also trick us, connect but also disperse us. It is perhaps not surprising that in 1890 when Edison designed and marketed a series of children's dolls that sang and recited nursery rhymes by the "magic" of a small wax cylinder, production lasted only six weeks. Not only did children find the dolls uncanny and difficult to operate, but the recordings very quickly disintegrated and became unplayable (Cowen, 2015).[1] Nowadays, electronic and digital treatments of the voice have brought into sharp relief its ontological plasticity – dramatizing the voice as an abundantly mediated object embedded in distinct socio-historical constellations. It is in this sense that popular music, with its overt manipulations, will be described in what follows as a "breaching experiment" in what voices can do. And yet while voices are shaped by specific socio-technical apparatuses, it is still important to know to what extent embodied singers, listeners and speakers learn to listen *through* these mediations and, indeed, gain pleasure from what mediation does. For as Neumark notes, "vocal

modalities…disturb, and thus make evident the *normal* operations of the voice" (Neumark, 2010: xv, original emphasis).

The starting premise of this article is that modern voices, particularly singing voices, are paradoxical: just as they hint at idealised notions of human talent, presence and expression, they also demonstrate sonorous acts to be radically relational in their co-evolution with a whole system of circuits and exchanges. Voices are compelling because they float between imaginaries of the unmediated and modern histories of mediation. While the privileged site of a hermeneutic density (the singing/speaking "I" as the sole locus of meaning), the voice, it will be argued, is best categorized as an assemblage in the sense suggested by Deleuzian theoretical currents and actor network theory: hybrid, multi-scalar, human–non-human events that emerge through nested constellations. After some theoretical preamble and ground-clearing, the article will briefly review how the voice's dalliance with its non-human "others" pours into it a tension between its supposed anchors in personhood and its hybridised construction: a tension that plays out visibly in the domain of popular music, where the pleasures of deconstruction of the singing voice are met with moral irruptions in response to its constantly evolving actualisations.

The final part of the paper explores how these ambiguities and extensions are expressed in relation to a single case, that of the Japanese virtual singer, performer and idol, Hatsune Miku. Miku is, in many respects, the contemporary equivalent of Edison's singing doll: a human-machine hybrid, circulating as merchandise, but who opens up pressing questions about where the voice sings from, where it resides, how it is heard and produced. Miku, it will be argued, is the inheritor of a cluster of vocal treatments in popular music, from the microphone to auto-tune, which demonstrate historical mediations at work. But she is also a uniquely globalised celebrity whose voice emerges as a wholly constituted object dependent on early 21st century developments at the intersection of crowd-sourcing, corporate investment and the increasing power of the algorithm. In this sense, Miku's voice is quintessentially assembled. But to assume that something as seemingly artificial works in total opposition to the "real" is untenable. Instead, the article will argue that Miku overtly stages the distributed ontology of all modern voices.

Mediations, Assemblages and the Post-Human Condition

To be interested in the multiple and cross-cutting ways in which musical and non-musical matter are interwoven is to be drawn to their constant interplay across time and space. The demands of inquiry invariably propel one on a journey through specific historical and regional cases, where the precise configuration of styles, practices and devices illustrate abundant fusions – of past and present, local and global, sounds and environments, bodies and works, flesh and silicon. Understanding these fusions is enriched by an engagement with strands of current theoretical and empirical work in the traditions of science and technology studies (Pinch and Trocco, 2002), cultural and media studies (Couldry, 2008), sociological and anthropological examinations of "musicking" (Small, 1998), as well as a whole raft of studies concerned with the "post-human" condition (Hayles, 1999). What many of these strands share is an attempt to make sense of the intimate foldings of agents, actions and objects, where it is no longer clear (if it ever was) where the boundaries between human and non-human lie (Haraway, 1991). For Hayles, as far as the realm of cybernetics and computer-mediated informationalism is concerned, ICT technologies have "become so entwined with the production of identity that [they] can no longer meaningfully be separated from the human subject" (Hayles, 1999: xiii); while for Haraway the image of the cyborg (a combination of the cybernetic and organic) is an "imaginative resource" (Haraway, 1991:

150) that not only captures our material reality but subverts boundaries that underpin regressive discursive alignments between woman and nature. Mol (2002) takes this line of inquiry further by showing how bodies and talk about bodies are not unified entities but multiple - realised in and through bundles of practices, practicalities and techniques at different sites. Hence, a disease is "what is done in practice" (Mol, 2002: 13), held together by diverse things in play: people, desks, chairs, textbooks, rooms, buildings, insurance, the computer system, the appointment letter, and so on.

To extrapolate from this work to the domain of music: sonic actors, forms and processes are inseparable from the cluster of entities which, through acts of perpetual mediation, shift the grounds upon which music is heard and sounds out. Here, mediation is a key term that refers us to a process of conducting one thing through another and the resulting effects of that conduction, the transformation of what is mediated. As a result of its mediation, an object's properties and contours change: it brings things into being. Hence, as Kember and Zylinska (2012) argue, it is just as important to pay attention to what emerges through processes of mediation as it is to raise the question of what is being mediated. This extends the idea of "media" as a discrete set of cultural forms and technologies that impact on us from a putative outside to the idea that such forms are always already entangled and embedded in the lifeworlds of audiences. Hence, "mediation is an intrinsic condition of being-in, and becoming-with, the technological world" (Kember and Zylinska, 2012: 1).

As far as music is concerned, Born's (2005) article "On Musical Mediation", is particularly useful in how it advances our thinking on music's changing ontologies. Drawing on the work of thinkers like Lydia Goehr and Alfred Gell, Born argues that mediation draws attention to the changing nature of the creation, performance and reception of music as they co-evolve with changes in and the availability of mechanical technologies. For Born, if we are to fully understand the way that music changes over time, how it coheres into different forms, then we need to understand the way that its ontologies are shaped by different mediating logics. To take one example, what music actually *is* changes as a result of its dissemination through myriad acts of mechanical recording and reproduction. With the gramophone and the radio, it's not just that a pre-existing body of sound is more efficiently disseminated as a result of a change in the background conditions. Rather, these conditions come to be experienced as embedded in the essence of what music is and what it does. Music's formal properties, its organization and its definition shift as well (Eisenberg, 2005). This is evident in the inseparability of modern popular music from its industrialization and mass dissemination, as Frith argues (Frith, 1988). With the act of mediation popular music is not just inflected by a change elsewhere in the system, but is intrinsic to and produces the world of popular music itself. When one adds to this the various historical trajectories of musical formations, inquiry can, as Born in a finessing of Adorno puts it, profit from adhering to a "constellatory conception of music's multiple mediations, understood…in the non-dialectical sense of the assemblage – of music's many simultaneous forms of existence" (Born, 2005: 13).

Here, the idea of the assemblage supplements mediation's emphasis on process with attention to the forms which provisionally congeal around music. In the traditions of Deleuzian philosophy (Deleuze and Guattari, 1981) and actor network-theory (Law, 1992), the assemblage is a non-reducible and decentered object which is constituted by an imbroglio of forces, favouring a "multiplicity of processes of becoming, affixing sociotechnical networks, hybrid collectivities and alternative topologies" (Farias and Bender, 2010: 2). While very much an open, ambiguous and fluid idea (a non-literal translation of the French term *agencement*, as originally used by Deleuze and Guattari), the assemblage concept offers up three interlinked propositions (Phillips, 2006).

Firstly, like mediation, assemblage thinking concentrates on how things are relayed *with* and *through* one another: it therefore links the problematic of structure with that of change, focusing on the dynamic character of the inter-relationships between the heterogeneous elements of any given phenomenon. While a kind of anti-structural concept, it still permits the researcher to register processes of "emergence, heterogeneity, the decentred and the ephemeral in nonetheless ordered social life" (Marcus and Saka, 2006: 101).

Secondly, in line with the post-human turn, assemblage thinking refuses to privilege human over non-human agency, instead seeing how they enmesh and activate one another. The premise here is the idea that material objects can engender action, one that avoids dismissing them as passive or instrumental tools that only come alive when they break down. Ascribing agency to things is not without its problems, of course: there are dangers inherent in assuming moral equivalence between the human and non-human worlds or to imply that they exist on symmetrical planes of meaning. Indeed, it is perhaps more accurate to say that agency is never the sole property of some imaginary unfettered individual directly casting their intentionality over the world of things. Agency is distributed, a result of our co-existence with the non-human world (Latour, 2005). It is in this sense that assemblages are proliferating hybrids that gather and are entangled with one another, garlands of actants that are interdependent and mutually co-shaping. Indeed, components of the assemblage only come to life in how they relate to other components, they have no interdependent life outside the network. So, although we act with conscious agency, we do so only in relation to a whole universe of others. We are not the pure authors of our actions, but these actions are sustained, deflected and activated in, with and by the non-human world.

One simple example may illustrate. Owners of new shoes often talk about "breaking their shoes in", by which they mean softening the shoes by using them over time until they become comfortable: the shoes yield straightforwardly to the imposition of human intentionality and action. Yet, is it not just as plausible to suggest that it is the shoes that are breaking us in? After all, it is surely the materiality of the shoes that inflames, then thickens, the user's skin to a point where they become comfortable? Or rather, it is in the exchange of properties between the shoe and the wearer that flesh and leather co-produce an eventual mutual adaptation that culminates in a wearable shoe. We can extend this to musical materialities, too, of course: the calloused fingertips of the guitarist can only be understood as a result of prolonged interchanges between fingers and strings which allow the instrument to become playable. The point is that musicking is not the sole capacity of an unfettered human musician whose instruments are mere puppets responding to an unbridled aesthetic vision: playing music (and, by implication, creativity) is always a relayed co-production.

The third component of assemblages relates to scale: an assemblage is multi-scalar and irreducible to either macro or micro levels, as if we could even separate these out: digital music consumption, for instance, is an assemblage that connects multiple levels at once, from legal and commercial processes to undersea cables, bodies and laptops (Prior, 2015). The local implies the global and vice versa—not just because of globalization and internetworked technologies that have telescoped information, time and space, but because the properties of the whole only emerge from the interaction between parts. Hence, holding to a hard and fast distinction between the local and the global is to short-circuit the analysis, as is the attempt to explain totalities or trace definite limits. Instead, the assemblage is, as Shaviro puts it, "an expansion of possibilities, an invention of new methods and new perspectives, an active 'entertainment' of things, feelings, ideas, and propositions that were previously unavailable to us". (Shaviro 2009, 148–9). As we will see, this expansion of possibilities is key to grasping contemporary vocal trajectories in the domain of popular music.

All of which moves us away from idealist notions of music as fixed in the score, or in the mind of the creative genius, towards a more radically relational approach which takes into account the chains of associations between human, institutions, technologies, texts, tools, instruments, works and so on. For Hennion, this makes the work of art, in general, an act of accumulated mediations, while music in particular "enables us to go beyond the description of technical and economic intermediaries as mere transformers of the musical relationship into commodities, and to do a positive analysis of all the human and material intermediaries of the 'performance' and 'consumption' of art, from gestures and bodies to stages and media" (Hennion, 2003: 84). For Born (2005), equally, a full and proper understanding of contemporary (particularly digital) music must recognise creativity as distributed or relayed between various productive capacities and forces that radically defer, de-center and de-hierarchise the work. In other words, music is an embedded object that emerges relationally in and through specific concrete acts and mediators with temporal indices. Or, as she puts it:

"…music derives from a continuous circuit of mediations and translations which demonstrate the mutable boundaries and connections between human bodies and subjectivities, scientific and visual representations, technologies and musical sounds" (Born, 2005: 25).

From the Grain of the Voice to Vocal Assemblages

If music works through the operation of the relay then so must its constituent sounds. But what does it mean to say that voices are mediated or assembled? What mediates what, and precisely how is it relayed and heard? There are three dimensions that I want to highlight in response to these questions.

Firstly, it's worth reiterating what is at stake in constructions and deconstructions of the voice: why does the voice, in contrast to the sounds of, say, the guitar, piano or drum, "poke out"? In the Western tradition, the voice certainly has a special place in systems of meaning and signification, an assumption made even more potent by the belief that the voice remains our richest form of communication and, therefore, an essential component of our humanity (Karpf, 2006). We talk of "giving people a voice" as an act of empowerment, for instance: we hear the voice as personally expressive, an index of someone's uniqueness. From a Derridean standpoint, speech is afforded special qualities in the Western tradition (Derrida, 1997). It is governed by a "metaphysics of presence", what Derrida calls "logocentrism", that positions the voice as closer to truth because of its proximity to the mind. Its status as vehicle for the communicative act also implies a constitutive role in what makes the social possible (Habermas, 1990), while biologists are wont to emphasise the mother's voice as the first thing a fetus hears and therefore the basis of a potent and lifelong social bond. This valorisation ramifies in music as the singing voice is central to a range of genres, transforming sounds into songs. In popular music it is placed in the center of the stereo field with the singer normally the focus of the audience's attention. Moreover, we often assume that vocal meaning emanates from the singer and expresses some*thing* about them – their personhood, feelings and experiences, for instance (Frith, 1996).

Secondly, because of these resonances, the voice has attracted a good deal of academic attention in sound, media and film studies, as well as in musicology (Neumark, Gibson and van Leeuwen, 2010). But academics still struggle to find precise vocabularies to convey the complexity of voices, including singing voices. Commentators have tended to fall back on psychoanalytic approaches (Dolar, 2006); or they have reduced the voice to the irreducible – to the singer's charisma, their personality or what Roland Barthes (1977) calls the "grain" of their voice, in which he assumes the form and physicality of the voice's originator - the body

- resides. Hence, the grain concept has been used by ethnomusicologists and sociologists to answer why "iconic" singers like Billy Holiday and Elvis Presley have such widespread appeal. For Potter (1998), for instance, Barthes' essay offers deep insights into the sensual dimensions of singing and sexuality because it excavates the encounter between language and voice; while for Frith, Barthes' essay is a useful starting point to explore how listeners intuitively sympathize with and gain pleasure from the voice as a "direct expression of the body" (Frith, 1996: 192). Markowitz, meanwhile, argues that Sinatra's vocal genius lay in the fact that he "embodied the grain of voice, and it embodied him. His voice did not express or reflect his life, personality, and world; it was the world around him" (Markowitz, 2009, n.p). In each case, the grain is a sonic externalization of an inner or corporeal truth.

Barthes' is a richly suggestive essay, for sure, but expecting it to retain leverage in a range of contemporary contexts stretches its credibility. Most importantly, the essay falls into the trap of positing an unadulterated self-presence, the assumption being that a holistic, pre-technological body engaged in direct communication is both possible and ideal—ideal because for Barthes, only certain singers are blessed with the rarity of the pure "grain". In contrast, for Sterne (2003), while histories of sound are also histories of the human body, that body is not something that is given prior to particular epistemic conditions. Just as hearing is not the domain of a transhistorical "pure interiority" so the voice only sounds out in a social space comprised of a whole panoply of discourses, techniques and machines that objectify and posit it as a particular kind of object and information. These include medical discoveries and audile techniques oriented to the mechanics of both the voice and the ear, like stethoscopes, hearing aids and headphones; understandings of how acousmatic sound works; and, of course, the advent of modern technologies of sound production and reproduction such as the radio, phonograph and telephone, that preside over massive changes in where the voice is, how it is heard and produced.

These devices didn't arrive without a good deal of shock and consternation, of course. As many authors have noted, they were accompanied by a heightened anxiety around the ghostliness of "disembodied" voices, where the ability to listen to and broadcast the voice without apparent corporeal substance was compounded by the fact that it could be preserved beyond death—literally, as a voice from beyond the grave (Sterne, 2003). Sound production devices appeared to many modern listeners to possess both too much agency and too little agency (where *is* the radio, after all?), interjecting between and damaging the purity of the face-to-face encounter. Indeed, for Connor (2001), the long technological effort to clean up the human voice and divest it of its hybrid traces (hiss, hum, metallic timbre, tininess) demonstrates the residual and paradoxical desire to reclaim it against the incursions of the machine. These paradoxes are still evident today. The moral outrage which often follows when singers are "caught out" lip synching and miming is part of this holding on to an imagined purity of talent and presence that is somehow bracketed off from the very conditions that made it possible to listen to these voices in the first place.

Thirdly, then, the singing voice is never a solo perceptual event, but is enshrouded in and mediated by a complex socio-technical and historically-specific set of conditions. It is an emergent outcome of complex relations that position it in specific ways across time and space. Paradoxically, while the voice attains its meaning as a uniquely expressive carrier in song, it is simultaneously accompanied by a whole machinic infrastructure (electricity, stages, acoustic treatments, amplifiers, microphones, compression and reverb units) which reveals that carrier to be radically hybridized. New ontologies are sparked into existence at the interchange of voice and modern recording technologies, human-machine conjunctions that recalibrate our conceptions about what and where a voice is and what it can do. As Frith and

others note, the voice of the crooner in the 1930s and '40s was sensed as an object of intimacy that could close the gap between voice and listener, just as it increased the distance between the voice and the singer's body. The effect was mediated intimacy, with affect being the currency of a new star system itself dependent on a series of novel body-machine techniques. After all, crooners did not just sing *into* the microphone but *for* the microphone (Penman, 2002). To re-evoke the co-evolving logics of shoes and strings mentioned above, crooners projected their voices into diverse spaces by literally bending their bodies and techniques towards the device—carefully policing distortion and avoiding plosives by moving in and out of the microphone's range. And around these vocal techniques amassed further specialisms, objects and discourses: pop shields, microphone stands, vocal booths, for sure, but also sound engineers—those unsung agents responsible for tending to the vocalist, while making aesthetic and technical decisions about signal chains, microphone positioning and so on that together shaped what the voice sounded like and how listeners listened (Kealy, 1979).

All of which favours a shift in our attention from *the* voice to vocal assemblages, comprising networks of vocality and myriad human–nonhuman imbroglios. Of course, there's a voice, with certain characteristics, timbres, cadences and tones: to deny this would be to deny any difference between singers whatsoever. It's just that there is nothing indivisible or solid here; the voice is never direct or primal. It emerges as a result of multiple associations and treatments. This doesn't mean that it is fake or inauthentic, merely that it is radically relational. It never speaks for itself but emerges, as Tiainen puts it, as "a processual factor that partakes in and initiates relational events where the states and capacities of the involved entities veritably transform—become anew" (Tiainen, 2015: 256). Indeed, it is precisely this emphasis on emergence and newness that allows us to pose two further questions: what are some of the key moments in the voice's hybridization and extension? And what role does popular music play in these new vocal becomings?

Vocal Breaches in Popular Music

At this point I want to briefly identify a number of modalities associated with modern histories of popular music and its emergence alongside voice-purposed techniques and technologies from the mid 20th century: synthesis, deconstruction and auto-correlation. An additional modality—simulation—will be the subject of the final section. Each comprises and helps give shape to a new type of assemblage, a new way that the voice is enrolled, ordered and extended. In playing with the boundaries around what is heard as "natural" and "normal", the four modalities parallel a distinct unease at the voice's capacities and possibilities, in turn necessitating a change in how voices are listened to. It is in this sense that popular music might be read as something akin to a set of breaching experiments as deployed by ethnomethodologists, a way of revealing the familiar, routine expectations of everyday life by disrupting them (Garfinkel, 1967). Paying close attention to what is being breached in expectations around the voice is therefore a particularly fruitful way to detect changing ontologies and conceptions of what vocal sound is and what it does. In what follows, the descriptions of these modalities will necessarily be short and suggestive, though fuller histories can be found elsewhere (see Dickinson, 2001; Biddle, 2004; Tompkins, 2010; Brøvig-Hanssen and Danielsen, 2016; Prior, 2018;).

Synthesis: firstly, with the invention and use of the Vocoder (voice-encoder), speech is turned into a packet of information that can be decomposed, transmitted and re-constituted. Designed as a military device to encode and encrypt vocal information, the Vocoder originally comprised a massive set of cabinets, capacitors and vacuum tubes. It was deployed

by Churchill and Roosevelt during the Second World War to communicate sensitive missives; while both Nixon and Reagan had Vocoders installed in their favored modes of transportation (Tompkins, 2010). By the 1960s, the Vocoder had been stripped back to become a more portable device that could make voices sound "robotic", and by the 1970s and '80s it was being used liberally by electronic groups like Kraftwerk, as well as electro-funk bands like Sun Ra, Afrika Bambaataa and the Jonzun Crew. Later, a keyboard was added, allowing musicians to "play" the voice like an instrument.

With the Vocoder, the voice is broken down and reconstructed and the result is arresting precisely because it appears to be enveloped in a not-quite-human presence. This is because the Vocoder voice is synthesized with a smooth space of flattened sonic frequencies and machinic fantasies. Its register is meshed with machines to produce a third entity—a cyborg voice that breaches certain expectations about where the human is or can be. According to Biddle, for instance, Kraftwerk's *modus operandi* has been to work against the "ego-centered phony aesthetic of presence and immediacy, articulated in particular by the vocal naturalism and emphasis on bodily 'performance' of the contemporaneous rock traditions" (Biddle, 2004: 82). Instead of rock-based ideas of presence, Kraftwerk's "cyborgian vocalities" herald the industrialised voice as yet another instrument, one that ironises, reflects and embraces an automated present. The synthesized voice is by definition a radically hybridised and defamiliarised voice. It therefore brings together on a symmetrical plane human and machine, to produce the conjunctive form "human-machine". Indeed, as Tompkins notes, the German version of "man and machine" is "Maschine-Mensch", "replacing conjunction with hyphen and allowing robots to make the band, one and the same" (Tompkins, 2010: 187).

Deconstruction: with sampling, on the other hand, every voice is open to being turned into digital information, every voice transmutable into data, lifted from its recorded context and recontextualised in a new one. This mirrors the phonograph's ability to move voices around without bodies, stripping them of their spatial and temporal context; but sampling twists this towards the digital as a set of binary logics that lay the voice open to myriad manipulations. Once digitally sampled, a voice becomes pure information, a series of 1s and 0s. This lends itself to speedy edits and global circulations: sampling musicians can make the kinds of intricate edits to voices that it would have taken months to achieve with tape splicing. Here, the sampler's interface is crucial because it represents the voice as a series of scrollable screens, menus and values that favour numerous and instantaneous rewritings. Filtering, bit reducing, time stretching, looping, chopping, pitch alteration, reversing, quantizing, mapping, tracking: these are just a few of the processes that are possible when the voice has become code, deconstructed and reconstructed through the silicon circuits and switches of the sampler. The list shows how malleable the voice becomes, how utterly breakable it is when deconstructed into bits. It also demonstrates the temporal and spatial dislocations that take place in the machine itself, as the voice is shifted around the grids of the sampler.

Early experimental uses of samplers like the Fairlight CMI and Emu Emulator demonstrate an emerging vocal aesthetic that was to become central to hip hop and electro pop: the voice extracted, chopped up, recycled and combined with a mélange of other sounds in a new sonic ensemble (Rodgers, 2003). From the punchy ejaculations of Kung Fu fighters to James Brown's "shouts", and from everyday street sounds to voiceovers for TV documentaries, sampled voices extend the musicians' sonic palette and their ability to juxtapose sounds in order to overlay new political meanings. This was an effect particularly associated with hip hop acts like Public Enemy, for instance; while Paul Hardcastle's 1985 hit "19" was distinctive for the quick-fire percussive effect of the main vocal line ("n-n-n-nineteen"),

digitally sampled from a documentary citing the average age of American soldiers in the Vietnam war. It's one thing to hear a voice "robotocised", quite another to hear it splintered into tiny sonic fragments that convey the abject horrors of war. By digitally de-anchoring the voice from its fleshy referents, the organic voice is replaced by its objectification as a uniquely pliable piece of data. A popular aesthetics of digital deconstruction, therefore, makes apparent and reinforces the voice as an informational entity—an entity that gives eventual rise to gratification among audiences who progressively get used to hearing the voice disassembled in this way.

Auto-correlation: pitch alteration is, at first sight, hardly the most distinct modality as far as processes of vocal mediation are concerned. Slowing or quickening the voice to increase or decrease the pitch was eminently possible long before the advent of digital samplers. In the 1960s, the high-pitched voices of the chipmunks in the Alvin and the Chipmunks animated music group were recorded at half normal tape speed, so that when the tape was played back at the regular speed it sounded a whole octave higher. Pitch correction has also long been a conventional treatment in popular music, one that often sits behind the scenes lest singers are revealed to be over-reliant on technological helpers. Indeed, the virtual studio plug-in, *auto-tune*, was designed precisely to subtly align vocal pitch without drawing attention to itself, a process that depends on an algorithmical treatment called auto-correlation. Auto-correlation is the correlation of a signal to a delayed copy of itself, the ultimate effect being that the software reads an incoming wave sample and moves it to the next nearest note, sharpening and flattening notes to a pre-set scale and at a rate that (at least the software's designers assumed) would not be too noticeable.

But the most audible and distinctive deployment of auto-correlation in popular music is its use in exaggerating the act of correction. When the retune value is set to an extreme rate, *auto-tune* draws attention to the voice as a hyper-melismatic machine, jumping instantaneously from one note to another, rather than sliding to that note. The result is some confusion as to where the identity of the singer and their voice resides, and therefore over the boundaries between organic and inorganic, real and unreal, and even male and female. Hence, with her 1998 hit "Believe", Cher's auto-tuned voice appears to be on the verge of breaking down as it leaps stepwise through the scale at strategic points in the song. According to Dickinson, the song unsettles naturalistic ideals of the female voice and its location in the unified body (Dickinson, 2001), because it reveals the organic referent to be an illusion, usurped by a cyborg version that collapses human and machinic imaginaries. Thus, Cher's voice "does not strike us as coming totally from within" (Dickinson, 2001: 226). For Brøvig-Hanssen and Danielson, the use of *auto-tune* among black American R&B artists like T-Pain, Frank Ocean and Rihanna, on the other hand, is one way the experience of being othered, alienated and dehumanized is expressed, while its widespread presence "represents a new and radical stage in the interaction of human and machine in the digital era of popular music history" (Brøvig-Hanssen and Danielson, 2016: 132).

Auto-tune is certainly one of popular music's most recognizable and well-used audio processing effects, a way of extending vocal capabilities in the soft spaces of digital audio workstations (DAWs). Yet its breaching qualities move in two directions: firstly, towards the distinct pleasures of hearing the voice hyper-mediated. This is evident in the spread and popularity of *auto-tune* smartphone apps and videos of newsreaders put through the software and uploaded to YouTube. Secondly, towards a critique of the software as a cheap gimmick or confidence trick, yet more evidence that pop music is fake and cosmetic.[2] In both cases, it is the fact that these desires and consternations index the voice as a constantly moving target

that is most important, with each mediation or set of mediations mobilising old and new discursive frameworks, positions and agents: some in favour of, and many set against, the implications of these mediations.

As for practices of listening, just as the voice is relayed through everyday circuits (radio, television, telephone, podcasting, video games, voice recognition software, voice over internet protocols, amplifiers), so these circuits have shaped what is heard as possible and normal. In other words, the ontological plasticity of the voice necessarily parallels changes in perceptual schemes involved in a whole historical filtering of listening. We hear *through* these mediations because the past inheres in our ability to experience the present and future. Milner (2009) alludes to this filtering when he asks readers of his book *Perfecting Sound Forever* to turn on the radio and listen to its voices, then turn it off and imagine what those voices would sound like were the people speaking in the same room as the listener. The point is that we listen to radio, and similarly mediated voices, only in relation to the accumulated experiences of the listener: the digital *through* the analogue, lo-fi *through* hi-fi, vinyl hiss *through* noise reduction, auto-tuned voices *through* radio voices and so on. The historically mutating voice therefore acts as a huge connecting force, an assemblage that subsumes multiple human and nonhuman voices across time and space.

Vocaloids and Virtuality: The Case of Hatsune Miku

In 2007 the Japanese software company Crypton Future Media in conjunction with the electronics company Yamaha released the second version of its Vocaloid software: an all-in-one software package that allows amateur and professional musicians to write songs without the need for a "real" singer. Based on the pre-recorded phonemes of a Japanese voice actress, the singing voice is drawn in with a mouse and keyboard in the Vocaloid score editor. Basic parameters such as pitch and note length can be edited, while users can select different vocal colorings such as "soft", "vivid", "sweet", "light" and "dark" to match the genre or mood of the song. The packaging mascot for the second version of the software featured a character called Hatsune Miku, an anime-styled 16 year old schoolgirl with long cyan pigtails. The idea being that Miku was the character whose distinctive high-pitched voice was being generated in this version of the software. It was her voice that users were co-producing, while Crypton's permissive legal framework (akin to a Creative Commons license) meant the legal strictures around ownership and copyright were relaxed: everybody could, in this sense, *be* Miku.

Within a few years, not only had fans generated a massive corpus of subsidiary works, such as fan art, that fed into and off Japan's developed *dōjin* system of amateur self-publishing, but something like 100,000 songs had been written. A good proportion of these were uploaded to Miku's official web portal, designated by Crypton as a digital repository of peer-produced songs. Meanwhile, a multi-media global ecology had sprung up around Miku, from digital games and YouTube videos to novels and karaoke packs. By 2009, Miku had gone "live" as an all-singing, all-dancing digital projection, fronting a series of concerts to sellout crowds in Asia and beyond. By 2014 she was performing on the American programme, The David Letterman Show, and in the same year opened for Lady Gaga on her "Artpop" tour. A fourteen date tour of North America was completed in 2016, followed by a collaborative art project at the Barbican Centre in London. One of her most popular songs, *World is Mine*, currently has 32 million views on YouTube, while the Vocaloid software has become part of the school music curriculum in Japan. In all cases, the songs designated as and performed by Miku are those written by her fans, the most popular of which are cherry-picked by Crypton from the online database to be performed "live".

It is against this background that some commentators have called Miku the world's "first crowd-sourced celebrity" and the "future of the music industry" (Hutchinson, 2014); while unsurprisingly, critics have baulked at the idea of a virtual pop star and deem her the ultimate fake celebrity. Academics, meanwhile, have attempted to take stock of Miku as a specific kind of collaborative media-mixed object. For Condry (2011), for instance, Vocaloid culture is essentially the expression of the creative energies of a fully-engaged community creating social value rather than economic value in the spaces of social media. In this view, Miku is the catalyst for a decentralised model of pop creativity, a horizontal media platform underwriting new possibilities for creativity amongst user-generated, participatory cultures. For Bell, on the other hand, Miku is a posthuman instrument that sits in a database paradigm of cultural production: a "means whereby something is achieved, performed, or furthered, and, especially, as a musical instrument" (Bell, 2016: 225).

In the context of this article, what interests me most about Miku, however, is her voice. For, at one level, while she is the latest in a long line of attempts to mechanically simulate the human voice, Miku is the ultimate post-modern singing machine. She never sings out of tune or forgets her words, she can potentially perform in many different venues at once, and, unlike her flesh-and-blood counterparts, will never be photographed *in flagrante*. In this sense, she is the ideal digital object, an avatar that can respond to the controlling fantasies of both *otaku* (obsessed male fans) and the *jimusho* system in general—that is, the tightly ordered model of signing, training and promoting idols in Japan. To accuse Miku of miming or lip-syncing would, after all, be absurd. As a hyper-real idol with a hyper-real voice she is already a profoundly hybridized human-machine. We are, in this sense, well beyond the modalities of vocal synthesis, sampling and auto-correlation: we have entered the logics of simulation and simulacra (Baudrillard, 1998). This doesn't mean that Miku's voice is non-material, merely that it exists in a different kind of materiality, one composed of diffuse digital bits spread exponentially through the circuits of corporate and peer-to-peer media. Her voice is a play of simulated surfaces across a graphic user interface and a projected digital image that refers to nothing but itself. Deconstructed and reconstructed in digital spaces, she represents the virtualization of the voice and its dislocation as it appears to drift further and further away from flesh and blood bodily referents.

But on another level, when one maps out the contributory forces that generate Miku's voice, one is left with the sense that it is the gathering point for various materials and forces and belongs to no-one and no-thing. Instead, her voice assembles at shifting points between the corporate and the collective, the underground and the hyper-commodified, the analogue and the digital. For instance, for all its virtuality Miku's voice still has a human referent in the form of Saki Fujita: the voice actress (the *saiyū*) whose voice provides the phonemes that constitute the Vocaloid sample database. It is Fujita's voice that is recorded, sold and "thrown" into the Vocaloid environment. Her "real" voice is sunk into a software environment that is itself a kind of virtual stage for further vocal manipulations. This means that despite the virtualization of Miku's body and voice, the chain of signification is still linked to the flesh, as Daniel Black (2012) argues. The body lingers even if the image is disembodied. Miku still gestures back to "real" voices and bodies: not only Fujita's, but those of vocaloid producers, the idealized youthful, hyper-feminine voice, the socially located bodies of fans, and the nurturing presence of the CEO of Crypton Future Media, Hiroyuki Ito. Indeed, while Hiroyuki is often seen as the "father" of Miku and in some respects (as a powerful corporate actor) speaks on her behalf, he is careful to acknowledge the lack of control he has over her as a distributed idol, a 21st century "read-write" object as opposed to a

20th century "read only" object. In that sense, he is one of a number of conduits for Miku's affordances, rather than her founding presence.

And lest we forget, Miku is an algorithm, a work of code. The software engineers at Crypton are constantly re-versioning her voice, upgrading it, re-coding it and improving it. The actualization of her voice is a work of constant maintenance, testing and research. One of Crypton's software engineers explained to me that one of his main jobs is to scour the Internet to find examples of new and experimental voices that might shape and inspire new versions of the Vocaloid software. He then takes test recordings with selected singers which takes a few days, spends another few days constructing a database of the sounds, and then brings all these materials into bespoke digital audio software to start the long process of analysing the sound, chopping it into little pieces, checking and modifying it. Once the database is complete, Crypton then decide on the different voice colourings for the character, which takes just under two months. They try out a few examples, employing musicians to write demo songs with the product; and also employ beta-testers to give feedback on the product outcomes. There are five people working just on the vocal samples at Crytpon, with another five on the software and a number of others on the online platform.

It is in this sense that Miku's voice is a "configuration of relationships among diverse sites and things" (Markus and Saka, 2006). It emerges in multi-scalar processes that fold together innumerably relayed materials. Hence, the terms "virtual" and "simulated" are only partially helpful here. "Assemblage" is better at capturing how Miku's voice is a convergent result of various human and non-human forces: silicon and carbon, corporate and grassroots, algorithmical and fleshy, local and infrastructural. Her voice is a channel, a database, a text and a massively collaborative accomplishment: it is a distributed force sustained through processes of circulation, a means of transmission and a repository of ongoing processes of participation. If people stopped composing her voice, it would wither precisely because it is brought into being, animated and activated in day-to-day practices.

As for the consumption side of things, listeners and audiences don't just get used to hearing the voice assembled in this way and find pleasure in these manipulations—"it doesn't take a human to sing a good song", as one young fan puts it (Verini, 2012, np)—but some participate in this very assemblage by producing Vocaloid songs in digital acts of ventriloquism and appropriation. By assembling her, they temporarily inhabit Miku, becoming part of the laboring force that constitutes her. Their participatory acts are constantly folded into the event, adding more socio-technical layers and proliferating temporal fusions into the Miku assemblage. On a number of occasions I've witnessed karaoke performers in Tokyo sing Miku songs. One of the most popular songs is *Miku no Shoushitsu*, ("The Disappearance of Hatsune Miku"). It's a song whose lyrics and video has Miku musing on her own digitality as a being delicately poised between absence and presence, a cluster of pixels that might get sent to the computer's trash bin at any minute.[3] At 240 beats per minute, the song is almost inhumanly possible for singers to keep up with, though many try. These are the entanglements of the assemblage on vivid display: flesh and blood karaoke singers in Japanese bars, affecting their voices to accompany the sounds of a synthesized voice designed to simulate singing voices, written by fans, shared online, performed by a digital avatar and circulated as a karaoke package. Unpicking these fusions is itself a job of some complexity, demanding methodologies fit for grasping the multi-scalar and emergent nature of things (Law, 1992).

Conclusion

This article has argued for a revised and reconstructed approach to the voice, one that expands our ways of knowing and hearing voices as assembled and interconnected. It draws on current thinking around the posthuman condition and assemblage thinking to demonstrate the heterogeneous and interrelated factors that constitute the voice and how we hear it. Even when confronted with something as seemingly primal, possessed and personal as the voice, what attention to assemblages demonstrates is the swarm of entities that are necessary for its sustenance in sonic worlds. This extends and radicalizes Derrida's deconstruction of the voice as the sole bearer of meaning, emotion and truth, but twists it towards the assemblage as an area of mediating conditions. Here, the voice cannot be reduced to the axiomatic of unadulterated human presence cut from the couplings and affective capacities generated by its engagement with myriad nonhuman others. The recent history of popular music demonstrates the ontological plasticity of the voice as it enjoins specific and embodied techniques, practices and devices. Attention to these temporary disruptions or "breaches" sharpens approaches to the distributed relationality of sound in general, and what Tiainen calls the "ever-evolving actualizations and reinvented possibilities [of the voice] in mediatized milieus" (Tiainen, 2015: 252).

The examples cited reveal the not-so-mute materials that sustain the voice, opening up important questions around the functions, states and capacities that transform its ontological grounds. This moves us away from the idea of technology as a contaminating obstruction to the "real" and towards the voice as a multiply-constituted object that emerges and re-emerges through the modulating forces of the socio-technical and material. Listening *through* these modulations allows us to examine the experience of listeners as filtered by histories of mediation, where sensory perceptions of what the voice is and can be are coloured by its bringing into being. That includes how technologies are appropriated and re-composed in ways that express the predicament of being a constructed "essence". We speak, sing and impute meaning to our utterances, but those meanings are always mediated by plural conditions and materials. But we also get sick, we affect, we impersonate, we age, and as we do so our voices and ears change and travel. If you've ever listened to your own voice as a mechanical recording, you might have experienced that uncanny sense that it is not entirely your own.

Is it really too much of a stretch, then, to say that all of our voices are distributed and composed? There are the obvious mediators: microphones, telephones, smartphones, radio, voicemail, gramophones and so on. But there are less obvious candidates like acoustic treatments in buildings that divide an outside from an inside, the dust and pollution that nestle within our respiratory systems, the water and alcohol that lubricate our throats, the food we eat, the medicine that deals with our illnesses and (perhaps) the tinnitus that transforms how we listen. Reconceptualising the voice, then, is not ultimately about hearing it as impure, artificial or mechanical, but about recovering it as a fully expanded, open system of co-evolving capacities. In short, as connective and connected.

References

Barthes, Roland (1977) *Image, Music, Text*, London: Flamingo.

Baudrillard, Jean (1988) *Selected Writings,* Stanford: Stanford University Press.

Bell, Sarah (2016) "The dB in the .db: Vocaloid Software as Posthuman Instrument", *Popular Music and Society*, 39(2): 222-240.

Biddle, Ian (2004) "Vox Electronica: Nostalgia, Irony and Cyborgian Vocalities in Kraftwerk's Radioaktivität and Autobahn", *Twentieth Century Music*, 1(1): 81-100.

Black, Daniel (2012) "The Virtual Idol: Producing and Consuming Digital Femininity", in Galbraith, Patrick W. and Karlin, Jason G. (eds) *Idols and Celebrity in Japanese Media Culture*, Basingstoke: Palgrave Macmillan: 209-228.

Born, Georgina (2005) "On Musical Mediation: Ontology, Technology and Creativity", *Twentieth Century Music*, 2 (1): 7-36.

Brøvig-Hanssen, Ragnhild and Danielsen, Anne (2016) *Digital Signatures: The Impact of Digitization on Popular Music Sound*, Cambridge, MA: MIT Press.

Condry, Ian (2011) "Post 3/11 Japan and the Radical Recontextualization of Value: Music, Social Media, and End-Around Strategies for Cultural Action", *International Journal of Japanese Sociology*, 20: 4-17.

Connor, Steven (2001) "The Decomposing Voice of Postmodern Music", *New Literary History*, 32: 467-483.

Couldry, Nick (2008) "Mediatization or Mediation? Alternative Understandings of the Emergent Space of Digital Storytelling", *New Media and Society*, 10 (3): 373-391.

Cowen, Ron (2015) "Ghostly Voices From Thomas Edison's Dolls Can Now Be Heard", *The New York Times*, May 5th, 2015: D2.

Deleuze, Gilles and Guattari, Félix (1981) "Rhizome", *Ideology and Consciousness*, 8: 49-71.

Deleuze, Gilles and Guattari, Félix (2004) *Anti-Oedipus: Capitalism and Schizophrenia*, London: Continuum.

Derrida, Jacques (1997) *Of Grammatology*, Baltimore: John Hopkins University Press.

Dickinson, Kay (2001) "'Believe'? Vocoders, Digitalised Female Identity and Camp", *Popular Music*, 20 (3): 333-347.

Dolar, Mladen (2006) *A Voice and Nothing More*, Cambridge, MA: MIT Press.

Eisenberg, Evan (2005) *The Recording Angel: Music, Records and Culture From Aristotle to Zappa*, New Haven, Conn. and London: Yale University Press.

Farias, Ignacio and Bender, Thomas (2010) "Introduction", in *Urban Assemblages: How Actor-Network Theory Changes Urban Studies*, Ignacio Farias and Thomas Bender (eds), London: Routledge: 1-24.

Frith, Simon (1988) “The Industrialisation of Music”, in *Music for Pleasure*, London: Routledge.

Frith, Simon (1996) *Performing Rites*, Oxford: Oxford University Press.

Garfinkel, Harold (1967) *Studies in Ethnomethodology*, Englewood Cliffs, NJ: Prentice Hall.

Habermas, Jurgen (1990) "Discourse Ethics: Notes on a Program of Philosophical Justification", *Moral Consciousness and Communicative Action*. Trans. Christian Lenhart and Shierry Weber Nicholson. Cambridge, MA: MIT Press.

Haraway, Donna (1991) *Simians, Cyborgs and Women: The Reinvention of Nature,* London and New York; Routledge.

Hayles, N. Katherine (1999) *How we Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, Chicago; London: University of Chicago Press.

Hennion, Antoine (2003) “Music and Mediation: Toward a New Sociology of Music”, in *The Cultural Study of Music: A Critical Introduction*, edited by Martin Clayton, Trevor Herbert and Richard Middleton, London: Routledge: 80-91.

Hutchinson, Kate (2014) “Hatsune Miku: Japan’s Holographic Pop Star Might Be the Future of Music”, *The Guardian*, Friday 5th December, 2014.

Karpf, Anne (2006) *The Human Voice: The Story of a Remarkable Talent*, London: Bloomsbury.

Kealy, Edward R. (1979) “From Craft to Art: The Case of Sound Mixers and Popular Music”, *Sociology of Work and Occupations* 6(1): 3–29.

Kember, Sarah and Zylinksa, Joanna (2012) “Mediation and the Vitality of Media”, *Life After New Media: Mediation as a Vital Process*, Cambridge, MA: MIT Press.

Latour, Bruno (2005) *Reassembling the Social: An Introduction to Actor Network Theory*, Oxford: Oxford University Press.

Law, John (1992) “Notes on the Theory of the Actor-Network: Ordering, Strategy and Heterogeneity”, *Systematic Practice and Action Research*, 5(4): 379-93.

Marcus, George E. and Saka, Erkan (2006) “Assemblage”, *Theory, Culture and Society*, 23(2–3): 101–109.

Markowitz, Robin (1998) “The Grain of ‘The Voice’”, http://www.culturalstudies.net/sinatra.html (accessed 19/1/16).

Milner, Greg (2009) *Perfecting Sound Forever*, London: Granta.

Mol, Anne-Marie (2002) *The Body Multiple: Ontology in Medical Practice*, Durham and London: Duke University Press.

Neumark, Norie (2010) “Introduction”, in *Voice: Vocal Aesthetics in Digital Arts and Media*, Neumark, Norie, Gibson, Ross and van Leeuwen, Theo (eds) Cambridge, MA and London: MIT Press: xv-xxxiii.

Neumark, Norie, Gibson, Ross and van Leeuwen, Theo (eds) (2010) *Voice: Vocal Aesthetics in Digital Arts and Media*, Cambridge, MA and London: MIT Press.

Penman, Ian (2002) "On the Mic: How Amplification Changed the Voice for Good", in Young, Rob (ed), *Undercurrents: The Hidden Wiring of Modern Music*, London: Continuum.

Phillips, John (2006) "*Agencement* / Assemblage", *Theory, Culture and Society*, 23(2-3): 108-109.

Pinch, Trevor and Trocco, Frank (2002) *Analog Days: The Invention and Impact of the Moog Synthesizer*, Cambridge, Mass.; London: Harvard University Press.

Potter, John (1998) *Vocal Authority: Singing Style and Ideology*, Cambridge: Cambridge University Press.

Prior, Nick (2015) "Beyond Napster: Popular Music and the Normal Internet", in *The Sage Handbook of Popular Music*, Andy Bennett and Steve Waksman (eds), London: Sage: 492-506.

Prior, Nick (2018) *Popular Music, Digital Technology and Society*, London: Sage.

Rodgers, Tara (2003) "On the Process and Aesthetics of Sampling in Electronic Music Production", *Organized Sound*, 8 (3): 313-320.

Shaviro, Steven (2009) *Without Criteria: Kant, Whitehead, Deleuze, and Aesthetics,* Cambridge, MA: MIT Press.

Small, Christopher (1998) *Musicking: The Meanings of Performing and Listening*, Hanover: University Press of New England.

Sterne, Jonathan (2003) *The Audible Past: Cultural Origins of Sound Reproduction*, Durham and London: Duke University Press.

Tiainen, Milla (2015) "Revisiting the Voice in Media as Medium: New Materialist Propositions", in *Media Matter: The Materiality of Media, Matter as Medium*, Bernd Herzogenrath (ed.), London: Bloomsbury.

Tompkins, Dave (2010) *How to Wreck a Nice Beach: The Vocoder From World War II to Hip-Hop*, New York: Melville.

Verini, James (2012) "How Virtual Pop Star Hatsune Miku Blew Up in Japan", *Wired,* 19th October, 2012: n.p.

Notes

[1] Though it is worth noting that nowadays those same wax cylinders can be played using a combination of microscopic and digital technologies that bypass the need for the needle to come into contact with the wax at all (Cowen, 2015).

[2] Two notes of caution. Firstly, none of these modalities are entirely discrete. Sampled voices are also synthesized voices, while changing the pitch of a voice is also a basic function of the sampler and tape manipulation before it. Auto-tune, meanwhile, is a sub-set of sampling, as are other algorithmical treatments like Melodyne, another piece of vocal editing software. Secondly, we have to be careful

not to assume that each wave of technology crudely determines the outcome: the relationship between music and technology is more subtle than this, involving affordance, misuses and breakdowns (see Prior, 2018).

[3] Indeed, we might speculate that it's precisely because Miku is an "imperfect" idol who lacks a logocentric core that fans feel drawn to compensate in acts of inexhaustible labour of love for a thing that might disappear at any moment.