

## 9 Digital Voices: Posthumanism and the Generation of Empathy

DAVID TRIPPETT

### Assistance

When Microsoft unleashed Tay ('thinking about you') onto Twitter on 23 March 2016, the AI bot progressively adapted to its environment, mimicking users and their memes in its 96,000+ tweets before it was taken offline. Modelled on a twenty-four-year-old American woman, Tay's life-span was just over fifteen hours. While the experiment in 'conversational understanding' famously ended with the bot issuing a hailstorm of racist, pornographic and offensive profanities (which Microsoft officially put down to users targeting its method of learning by imitation), Tay also produced some weirdly unparroted responses. To the question: 'Is Ricky Gervais an atheist?' she answered: 'ricky gervais learned totalitarianism from adolf hitler, the inventor of atheism'.<sup>1</sup> No information was released publicly about the algorithmic route to such deductions, and it would seem the press verdict rang true: 'this is an example of artificial intelligence at its very worst – and it's only the beginning.'<sup>2</sup> The interface for her offensive utterances was silent, the flickering screen whose electronic text can stimulate verbal sounds only in the reader's mind. But an interface is inherently reciprocal: a gateway that opens up and allows passage to some place beyond (Galloway 2017, 30). Imagine if Tay could talk out loud. Imagine the sound of her voice, her persona-in-sound.

In an age of digital culture, where seemingly all aspects of the aesthetic experience of sound are soluble in information and communications technology, the relations between human and smart device remain both controversial and banal: controversial in objectifying aspects of human interactivity; banal in instilling a 'passion of the object' and its agency (Baudrillard 1988, 84). The rich history of debates on this topic takes a particular turn with the advent of digital processing. While many commentators have celebrated the affordances of digitised music and musicianship, others have sounded the alarm: 'composition, musical praxis and musical perception sit at a crossroads. The rapid development of the digital world together with their networking will not remain without consequence for musical creation' (Kriedler et al. 2010, 8). Another decade into the twenty-first century, and consequences are not hard to identify: the

[227]

co-dependence between new digital media and music is so extensive that it is hard to exclude any areas of musical praxis from the affordances of new technologies. From algorithmic playlists and virtual mixing desks to in-ear speech processors and the infinite potential of (sensory) signal transduction, digital media enable and characterise virtually all access to and production of mediated sound.

This chapter looks critically at the ramifications of digital technologies that variously assist, enable or simulate musical praxis. A brief overview of the steps ahead may be helpful at the outset. The first section sets up an opposition between the idea of the digital *tool* that expands or augments human agency, and the machinic automatism predicated on the idea that reality is fundamentally number (dataism) – with matter as its mere vehicle – and ticks along mechanically without the need for human consciousness. This move gives rise to the idea that quantifiable, mechanical automatism is also intrinsic to human agency, a strand of posthuman thought on which the rest of the chapter turns. Accordingly, the second section shows how algorithmic composition may be posed as an expression of the posthuman, but this rapidly becomes untenable: because of its reliance on human evaluation, algorithmic composition can be better explained in terms of the tool/augmented humanity model rooted in ‘assistance’ and its processes of collaboration and interaction. This essentially reassuring conclusion comes under increasing pressure in the final section, which focuses on the synthetic voices of digital assistants – Siri, Alexa et al. – from online service providers. Here the generation of empathy appears to satisfy practical requirements (e.g. care for the elderly or increasing the desire to buy products) but, in fact, in the context of conversation in particular and interaction in general, it can pose problematical issues of a surrogate ‘conscience’. Accommodating this within a humanistic model is still possible, but a closing case study of Tod Machover’s futurist opera, *Death and the Powers* (2010), raises the prospect of what might be called a ‘dark ontology’ of the digital that can’t easily be shrugged off.

Whether explicitly or not, ‘assisting’ technologies necessarily establish a relational dynamic with users. In a sense, music technology in its machinic guise has always aided or assisted people at work, whether figured as mechanical device or software. For Lewis Mumford, writing in 1934, the machine’s justification lay in eliminating human slavery or its modern equivalent, the need for people to carry out menial tasks. Or indeed, any work at all: ‘the ritual of leisure will replace the ritual of work,’ he predicted, ‘and work itself will become a kind of game’ (Mumford 2010, 279). On the one hand, music notation software undoubtedly reduces the

labour of writing out or otherwise preparing individual parts, to cite an immediate example, and digital sound editing renders the manual cutting and splicing of magnetic tape a comfortable metaphor. But on the other hand, Mumford's prediction can sound naïvely optimistic – perhaps only possible amid the heyday of Fordism and workplace automation – when faced with the cognitive operations of skilful, creative activities such as composition and performance. To regard such operations as susceptible to machinic 'take-over' is to injure our sense of what is special to traditional music-making. Here, Mumford's distinction between *machine* and *tool* becomes helpful:

[T]he tool lends itself to manipulation, the machine to automatic action . . . The difference . . . lies primarily in the degree of automatism they have reached: the skilled tool-user becomes more accurate and more automatic, in short, more mechanical, as his originally voluntary motions settle down into reflexes, and on the other hand, even in the most completely automatic machine, there must intervene somewhere, at the beginning and the end of the process . . . the conscious participation of a human agent. (Mumford 2010, 10)

In other words there is a continuum between agency and automation in the use of tools to enhance 'human' creativity. As we shall see, this principle applies to the digital manipulation of sound as much as the Model T assembly line.

Moving into the specific discourse of sound technology, it is not uncommon to speak of digital 'tools'. Digital recorders are essential tools for field recordists just as the digital camera is a 'generic tool' for students of visual culture (King et al. 2016, 141). Plug-ins for audio editing are tools that extend an editor's capabilities just as robust direct drive turntables permit the scratching and backspinning necessary for turntablists to turn a record player into a musical instrument. All imply a dependent relation between user and tool. In these examples, a 'tool' appears as a kind of neutral enabler, a means to an end, seemingly more than mere equipment yet less than prosthesis (where 'prosthesis' entails an addition or extension 'of ourselves' in both physical and virtual contexts). But when data, the basic substance of communication itself, become the enabler of activity, any manipulation of such a tool becomes reflexive and correspondingly powerful as an agent that determines outcomes. 'Data itself is a tool', writes the former director of Google Ideas and former CEO of Google, 'and in places where unreliable statistics about health, education, economics, and the population's needs have stalled growth and development, the chance to gather data effectively is a game-changer' (Schmidt and Cohen 2013, 15). Ostensibly sound data can't easily be utilised in this way because they do not represent other things; they simply are what they are, though they can

readily be illustrative of prior content, such as the sonification of medical imaging (when converting the visual data of a brain scan into sound, for example). So we might say that sound data relate to digital ‘tools’ as numbers relate to mathematical formulae: their manipulation is determined by the shape and function of the technological means available to the manipulator more than by notions of creative intent. In other words, while we are familiar with tools that enhance our own agency, as tools become increasingly information-driven they begin to take on attributes of agency in their own right.

One practical outcome of the digital organisation of sound through binary code is its enhanced potential for manipulation (recall this is Mumford’s principal criterion for identifying ‘tools’), both manually by people coding and by algorithm:

once coded numerically, the input data in a digital media production can immediately be subjected to the mathematical processes of addition, subtraction, multiplication and division through algorithms contained within software. (Lister et al. 2009, 18)

Of course, manipulation of sound through software and physical manipulation of, say, a piano keyboard are not of the same order. One deals in layers of matterless signs, the other in tactile resonance. Yet the very point of difference – binary code – forms the tell-tale thread for theorists such as Lev Manovich to illustrate how the histories of computing and audio-visual media are entwined. The individual static photographs of the cinematograph store audio-visual data, just as do the 1s and 0s of an MP3 file, itself a product of computing: ‘this is why the Universal Turing Machine [the first stored-programme computer] looks like a film projector’ (Manovich 2001, 24). Both store data according to the split-second pauses of an on/off regularity; one on celluloid, the other electronically. As one commentator put it, for Manovich ‘the flicker of film was already a digital flicker’ (Galloway 2017, 4).

This helps explain why, for Manovich, numerical coding constitutes ‘the new universal intermediary’ that bridges senses and audio-visual media, whether via voltage fluctuation and the pixels of LCD displays or the travelling waves of oscillating pressure that we experience as sound (Manovich 2013, 153). In this view digital media are constituted by the data flows in which all input is converted into numbers. For music historians, then, digital media arguably have a genetic relation to early Greek understandings of music: for the Greeks musical sound points to a background numerical order, whereas with digital media an underlying numerical order gives rise to musical sound (this is the dominant model, but both processes are ultimately reversible). In this sense, Pythagoras’s

interpretation of music as sounding number begins the circle that digital audio closes, for – materially speaking – digital audio is nothing but its binary data (quantities of sound coded into a cultural form) while Pythagoreans believed sound was essentially the representation of a numerical natural order, a cosmic worldview ensounded through the ratios of resonating intervals. Both constitute processes of abstraction into the symbolic realm of mathematics, but with opposite directions of travel: from sound to number; from bits to sound.

The idea of ‘dataism’ underpinning this view is broader than musical discourse, of course, and is worth explaining briefly.<sup>3</sup> It interprets the world as consisting of data flows, and values any technology – or entity more broadly – according to its contribution to data-processing. Some futurists have interpreted this as offering the promise of an emerging universal language, a technological tower of Babel to unify all disciplines: ‘According to Dataism, *King Lear* and the flu virus are just two patterns of dataflow that can be analysed using the same basic concepts and tools’, writes Yuval Harari briskly (2016, 367–8). Hence the simplifying, optimistic prognosis that follows: ‘This idea ... gives all scientists a common language, builds bridges over academic rifts ... Musicologists, political scientists and cell biologists can finally understand each other.’ Is interdisciplinarity to become a tautology? I suspect that this model of binary code’s infinite transferability would be hard to replicate in units of data equally comprehensible to all academic disciplines.

How exactly do digital musical tools enter this story of digital morphology? Course reading lists and university library holdings offer empirical evidence that the aural dimension of data has been disproportionately subordinated to the visual dimension within media studies and the philosophy of technology. One possible reason is that a progressive dematerialisation of the image was accompanied by a radical materialisation of sound prior to digital recording. As Peter Weibel notes, image capture proceeded from the daguerreotype, which freed the artist’s hands, to telegraphy, electronic scanning and film – all of which worked through time – and on to video: ‘the basic conditions for electronic image production and transfer’ were established through a combination of the electron, magnetic recording and cathode-ray tubes, and thereafter ‘matterless signs’ – enabled by integrated circuits, transistors and silicon chips – came to drive a ‘post-industrial telematics culture’ (Weibel 1996, 338–9). Each stage in this historical model for image production and transmission offers a further abstraction from real-world visual experience. Sound recording went in the opposite direction. After centuries of music notation based on abstract, legible signs, successive technologies objectified the physical matter of sound vibration, writing the impression of sound waves onto smoke-blackened paper, into

wax, shellac, acetate and then vinyl discs, and finally reconfiguring it as forces within a magnetic field. Such writing of sound was no longer symbolic but material. This radical process of materialisation only ended around 1980 with digital sound recording, which makes the abstraction of sound data 150 years younger than that of visual data.

Regardless of their relative youth, digital musical technologies have raised searching questions about human relations to music. We might helpfully consider these in light of the distinction between tools and machines discussed above. ‘Digital technologies can be used as musical tools’, Andrew Brown concludes straightforwardly. ‘Just as an audio amplifier can make music louder, a music technology can be an amplifier of one’s musicianship, enhancing musical skills and increasing musical intelligence’ (2015, 6). But while we are comfortable with the idea of digital technology assisting us (remember the springy paperclip in MS Word?), cognitive enhancement is of a different order because it concerns interior thought, whose secrecy and non-transferability had been regarded as inviolable under the tenets of a liberal humanism. That status changes the moment human ability, skill and intelligence are ‘amplified’ by digital technology.

This reflects larger discourses on artificial intelligence (AI). A common question for students of AI and prosthesis goes as follows: if a blind man walks with a cane, is the cane part of the man? Ostensibly, the cane is a tool, separate to the man. Conceived as a single unit, however, the man + cane function is understood as a self-correcting process where the cane becomes a sensory extension under his control (hence a closed system with feedback loop, cybernetically speaking). The philosophical argument for extending the mind by expanding the domain of cognitive *control* was first advanced by Andy Clark and David Chalmers:

If . . . a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* part of the cognitive process. (Clark and Chalmers 1998, 8)

This has implications for generative software, including certain models of algorithmic composition; it gave rise to the concept of distributed cognition wherein human agency and thought are ‘enmeshed within larger networks that extend beyond the desktop computer into the environment’ (Hayles 2012, 3).

Keyboard players and other instrumentalists have long been familiar with the notion that a resonating instrument comes to feel as an extension of the body (de Souza 2017; Le Guin 2006); for unamplified singers, the body *is* the instrument, collapsing any distinction between instrumentality and self. The same relation obtains for scholars instantly accessing vast

data networks via the computer keyboard, with the difference that while a musical instrument is physically finite, online networks grow exponentially, supported by ever-larger servers. When living in a small town, Hayles explains, the effects of this hand–keyboard compound only become apparent when we lose the networked extensions we take for granted: ‘when my computer goes down or my Internet connection fails, I feel lost, disorientated, unable to work – in fact, I feel as if my hands have been amputated’. Put another way, the embedding of computational media into the environment is having ‘significant neurological consequences’ (Hayles 2012, 2, 11). The physical breach – components extending and rearticulating our ‘selves’ in McLuhanesque ways – marks a threshold in self-perception and the use of tools. It is also the linchpin around which posthuman discourse has come to structure a future-orientated understanding of human relations to digital technology. In most encounters with data, then, what might be seen as an abstraction to number is better seen as an expansion or augmentation of the human body-in-action.

### **Mimesis: Creating with Algorithms**

Having touched on ‘prosthetic’ additions to the body and the implied tension between literal limbs and metaphorical uses of the concept (Sobchack 2006), it might be useful here to clarify three distinct terms that pertain to human-technology relations. In brief:

**Transhumanism (H+):** a futurist-orientated intellectual and cultural impulse that believes in enhancing the human condition through technology in all its forms. The result is an iteration of *homo sapiens* enhanced or augmented, but still fundamentally human. A central premise of transhumanism is that biological evolution will eventually be overtaken by advances in genetic, wearable and implantable technologies that artificially expedite the evolutionary process (More 2013, 3–8).

**Posthuman:** the condition attained after stages of technological enhancement of the human condition render the subject no longer normatively ‘human’. In its more radical iterations, this condition does away with the biological body altogether, where information patterns uploaded to a fantastical supercomputer suffice to constitute a posthuman identity. In an extreme form of noetics, such existence becomes all mind, more powerful than present minds, employing ‘different cognitive architectures or includ[ing] new sensory modalities’ (Bostrom 2009, 347).

**Posthumanism:** a discursive web of philosophical positions defined against, and seeking to supplant, the autonomous liberal human subject and its concomitant anthropocentric bias, excessive valuation of human achievements, and preoccupation with humanity's supposed differences from (and superiority to) the rest of animate life. This asserts no definite break between sentient and non-sentient matter in the relational fields of an environment because matter is no longer conceived as passive or inert, but capable of 'self-transformation, self-organization, and directedness' (Coole and Frost 2010, 10). Posthumanist attitudes anticipate an increasing incorporation of artificial technologies into the body not primarily as enhancement of the human condition (as in transhumanism), but as its anticipated dissolution: this is seen as part of a more fundamental dissolution of literal boundaries between subject and object, body and environment, and a corresponding recalibration of our sense of self-identity within a world of objects. The result is an identity defined more by its controllable architecture rather than its cultural history.

The first two terms denote subject positions in relation to technological objects, the third an attitude that encompasses these as part of a broader critique of the humanities. All intersect with the idea of computer 'assistance' when that assistance exceeds simple tasks such as calculating quotients or formatting scores ('task intelligence'). But when a device's intelligent behaviour appears to be purposive, it is the posthuman that is most openly implicated.

In order to see these categories in action, consider algorithmic composition, perhaps the most tangible instance of human utilisation of music's mathematical undergirding. Within the contemporary scene of melodic composition, such tools as Dirk-Jan Povel's *Melody Generator* (Povel 2010) or Dmitri Kartofelev and Jüri Engelbrecht's 'structured spontaneity'<sup>4</sup> use Markov chains or fractal geometry to create melodies based on restrictive programming of overtone properties and predefined stylistic parameters. Each melodic pitch and rhythm is calculated from a mapping of the immediately preceding pitches and rhythms in a theoretically endless linear operation. This approach to melodic data mirrors statistical analysis of the same (e.g. how many sixteen-bar melodies from 1790 have forty-eight notes?), and here the role of cognition recedes, including – crucially – the responsibility to assess the resulting music. In substituting for, and calculating faster than, human cognition this approach embodies a posthuman perspective, and as such it is fundamentally different from the transhumanist perspective embodied in the established canon of



algorithmically composed music that fulfils the criteria of originality and aesthetic appeal. This ranges from Charles Dodge and Clarence Barlow to Herbert Brün, G. M. König, and Iannis Xenakis, whose *Gendy3* (1991) typifies the composer's lifelong pursuit of 'automated art': art that draws on the processing power of a computer to extrapolate the implications of artistic thought, thereby still 'reserving for the human the role of creative decision making', and accordingly posing challenges for any musicological analysis that seeks to 'reconstruct the laboratory conditions of the algorithmic creation process' (Hoffmann 2010, 121, 129).

While machines do have feedback loops, the value of these is a matter of dispute. Could they ever equate to a kind of music criticism? As Christopher Ariza notes, music generation as a closed cybernetic circuit cannot be regarded as composition proper, for it amounts to a game of data manipulation, whose results are barely evaluated by a machine and become meaningless when presented to human listeners who try to distinguish machine melody from human melody: 'use of the [Turing test] in the evaluation of generative music systems is superfluous and potentially misleading; its evocation is an appeal to a measure of some form of artificial thought, yet, in the context of music, it provides no more than a listener survey' (Ariza 2009, 49). In such a view, human and machine listening remain radically asymmetrical.

Wrapped up in the irreducibly human perceptual character of music assessment is the broader issue of consciousness, one strand of which is whether the brain thinks and feels only, or also calculates. Would a computer engaging in a process of generating music experience release on the final downbeat, exhilaration at the rhythmic vigour, or a sense of progressive harmonic tension and release? How would we know? Concomitantly, at what point would a court need to recognise the computer as owner of an intellectual property arising from a musical sensibility, that is, as an autonomous creator?

This debate has played out in contexts far beyond digital music, of course. A summary of perspectives – alluding to everything from a Turing test to David Chalmers's 'psychophysical principles' (1995) and 'information that has lost its body' (Hayles 1999) – is given in the Socratic dialogue from Tom Stoppard's 2014 play *The Hard Problem* (whose title borrows Chalmers's own term for scrutinising consciousness). Here, post-docs at a prestigious institute for brain science argue over the nature of consciousness within a calculating machine:

AMAL: Sure, but the brain is a machine, a biological machine, and it thinks. It happens to be made of living cells but it would make no difference if the machine was made of electronic gates and circuits, or paperclips and rubber bands for that matter. It just has to be able to compute.

- ...
- LEO: What [a computer is] doing is a lot of binary operations following the rules of its programming.
- AMAL: So is a brain.
- LEO: But can a computer do what a brain can do?
- AMAL: Are you kidding? – A brain doesn't come close!
- ...
- HILARY: It's not deep. If that's thinking. An adding machine on speed. A two-way switch with a memory. Why *wouldn't* it play chess? But when it's me to move, is the computer *thoughtful* or is it sitting there like a toaster? It's sitting there like a toaster.
- LEO: So, what would be your idea of deep?
- HILARY: A computer that minds losing.

(Stoppard 2015, 22–3)

In a gendered division, the project leader, Leo, mediates between Amal's futurist agenda and Hilary's empathy and subsequent concession that it would be impossible to tell whether a computer minds losing or not. Unlike Socratic methods, however, there is no resolution free from contradiction, which reinforces the conundrum: the unknowability of a computer's aesthetic judgment upon experiencing its own creation.

Deterministic algorithms will always produce the same output from a given input. For creative composition in any genre this intrinsic limitation is undesirable in isolation, even as computation can explore configurations of pitch and rhythm unavailable to humans alone. 'A composer who knows exactly what he wants, wants only what he knows – and that is one way or another too little' explains Helmut Lachenmann (Lachenmann and Ryan 1999, 24). Not all algorithmic approaches to music generation work in a directly automatic way, of course, and there are a range of interactive processes that do not require real-time human input (Nierhaus 2009 offers a helpful overview). To remain with melody generation a moment longer, just as the nineteenth-century critic Eduard Kulke – inspired by Darwin and Lamarck – believed melodies were subject to evolutionary principles, and proposed genealogies of melodic transformations as part of a collective cultural memory (Kulke 1884), so Francesco Vico's computer system Iamus (2010) uses a 'genetic algorithm' that mimics the process of natural selection. It generates random musical fragments, mutates them and determines whether they conform to pre-defined rules (genre-specific, instrument-specific, stylistic). By this process, all fragments are incrementally refined into rule-adhering music. In the domain of rhythm, Eduardo Miranda's *Evolve* likewise uses the interaction between multiple algorithmic generators to compose repertoires of rhythms: 'the agents were programmed to create and play rhythmic

sequences, listen to each other's sequences and perform operations on those sequences, according to an algorithm' (Miranda 2012, 225).

Under conditions of improvisation, by contrast, such a process could not be entirely automated, of course, for human input is needed in real time. Here the distinction between the roles of computational creativity and machine assistance is significant. While evaluation criteria cannot be clearly stated in a programming language, Interactive Evolutionary Computation allows for interaction between the algorithm and human participants, suggesting that what is often put forward as computational creativity is in reality better understood as augmented human creativity according to the tool model given above, or as collaborative interaction. One example is John Biles's jazz melody generator GenJam, described as 'a genetic algorithm-based model of a novice jazz musician learning to improvise', in which a human mentor gives real-time feedback which is then absorbed by the programme to improve the future generation of melodic patterns, i.e. in a closed-loop feedback function (Biles 2007, 137). This offers a model of collaboration between human and computer, where the dominant operator is the computer: the digital element relies on seemingly unprogrammable human decisions, making it *human-aided* rather than *computer-aided* composition. By the same token, George Lewis's Voyager software – a 'nonhierarchical, interactive musical environment that privileges improvisation' – offers an instance of man–machine collaboration on more equal terms. The programme's analysis in real time of a human improviser both guides its own independent behaviour arising from pre-defined algorithmic values, and generates 'complex responses' to the human musician's playing. 'Voyager is not asking whether machines exhibit personality or identity,' Lewis remarks, 'but how personalities and identities become articulated through sonic behavior' (2000, 38). Such software, defined by interactivity and response, is positioned as a tool for exploring the creation of values, our own creativity and intelligence as well as 'musicality itself'. It is as much an epistemological tool as a compositional agent.

### **Semiosis: Sounds like Human**

*Sound* tools engage the discourse on digital assistance perhaps most directly through simulations of the human voice, aka speech synthesis. This personalises the interaction with a computer intelligence. Synthetic voices without a body instinctively conjure up a hybrid persona, a virtual personality, touching on a network of signs that distinguish human from non-human, a cultural and biological distinction that has come under

pressure in recent years (Clark 2004; Bennett 2010; Bostrom 2014). While it is not hard aurally to distinguish human voices from robotic alternatives, the current reliance on the data networks of female-sounding digital assistants – Siri, Alexa, DeepMind, etc. – verbally responding to our commands offers a sonic analogue to Hayles’s hands that feel ever more part of the computer she controls, or composers who make use of borrowed processing power to uncover the ramifications of a melodic cell, a rhythmic pattern or a stochastic principle.

From the perspective of linguistics, it was the advent of phonology in the early nineteenth century, with its differential oppositions, that allowed for a computational approach to speech sounds in the twentieth. Chains of phonemes, those ‘senseless atoms that, in combination, make sense’, as Mladen Dolar put it, could now be organised by the oppositional logic of binary code:

All the sounds of a language could [now] be described in a purely logical way; they could be placed into a logical table based simply on the presence of absence of minimal distinctive features, ruled entirely by one elementary key, *the binary code*. In this way, most of the oppositions of traditional phonetics could eventually be reproduced (voiced/voiceless, nasal/oral, compact/diffuse, grave/acute, labial/dental, and so on), but all those were now re-created as functions of logical oppositions, the conceptual deduction of the empirical, not as an empirical description of sounds found. (Dolar 2006, 19)

In like fashion, digital algorithmic composition can be thought of as a cultural technique in Bernhard Siegert’s terms of a self-referential symbolic practice in between object and sign, a practice whose ‘operative chains . . . precede the media concepts they generate’ (Siegert 2015, 11). Music-making exists without (and before) explicit music theory, and likewise music-calculating exists before music-computing: ‘people wrote long before they conceptualized writing or alphabets . . . Counting, too, is older than the notion of numbers’ (Macho 2003, 179; cf. Siegert 2015, 11). In contrast, the simulation of personal voices fits into a sub-category of Anglo-American posthumanism, namely, representations of human identity arising from chains of operations that produce something that appears to be real, but is in fact only its semblance. An example would be that when absorbed in a crime drama you don’t notice the TV as a medium unless the signal malfunctions. Media like to conceal themselves, in other words, which, for digitally synthesised voices, means creating the auditory semblance of an entity that is actually talking to you.

The formulaic sentences of commercial chatbots quickly betray the limited repertory of a non-adaptive mind. Once digital tools assume a persona, by listening to speech commands and uttering informed (if not

‘thoughtful’) responses, they become virtual ‘assistants’ in a more than metaphorical sense. Such technologies, dating back to IBM’s Shoebox in 1961, are now commonplace, arising from leading commercially funded research and development budgets at Apple (Siri), Amazon (Alexa), Microsoft (Cortana) and Google (DeepMind), each vying for market dominance. But even while AI offers different (computational or information sourcing) capacities to human cognition, developers’ ambitions point directly towards the goal of sounding like humans because the creation of empathy sits at the heart of the project of an artificial simulacrum: ‘I truly believe that for AI to be useful in our daily lives, it has to be something you can connect with. Conversation is the next step, to be more human-like’, explains Rohit Prasad, Amazon’s chief scientist for Alexa. ‘We could cross the 10-minute barrier now, but 20 minutes is extremely hard. This will be a long journey.’ In fact, the university-based ‘Alexa Prize’ (\$2.5 million), inaugurated in 2017, is stimulating research in precisely this field; users utter the command: ‘Alexa, let’s chat’ to sample one of the current loquacious ‘socialbots’ under development. ‘But there’s also dynamics here where you want the AI itself to come back and trigger some conversational topics with you’, Prasad explains. ‘That it can tell you “Hey I can talk about . . . Did you hear about this event?” for instance.’<sup>5</sup> Information-based responses seem set to continue, in other words, rather than interaction resembling conversation proper, ‘the great paramount purpose of social meetings’ as Thomas de Quincey (1863, 151) famously put it.

Since these devices don’t have a freely active learning function like Microsoft’s Tay, their automatic ‘jokes’ only elicit groans: ‘Want to hear a dirty joke?’ asks Alexa. ‘A boy fell in some mud. How about a clean joke? He had a shower.’<sup>6</sup> For now, then, empathy generation appears most scrutable through the code its programmers use to modulate vocal intonation and timing: the sounding voice. By necessity this is declarative, specifying whispers, bleeped-out expletives, speech delivery and even substitute words in crudely literal ways. To code a whispered sweet nothing in Amazon’s Speech Synthesis Markup Language (SSML), it’s `<amazon:effect name=“whispered”>`; to programme an expletive bleeped out, it’s `<say-as interpret-as=“expletive”>`; the ‘prosody’ setting meanwhile permits alternations of tempo (‘fast’/‘slow’) and pitch (‘low’/‘medium’/‘high’) to resemble an expressive mind behind the synthetic phonemes. And programmers have been swift to exploit the capacity for irony. Freia Lisa Lobo’s split-personality quip is a case in point: “[blank female voice] Right now in New York it is 70 degrees. [Pause. Whispering] I see dead people’.<sup>7</sup>

Colloquialisms or ‘speechcons’ have been available since February 2017, crudely generating empathy with interjections like *cheerio*, *argh*,

*d'oh*, as well as *booyah* and *bazinga*, to humanise the monotone obedience of the prim robo-voice. By prefacing these utterances with a 'say-as' tag, the software understands they are to be emphasised in a further step towards the acoustic semblance of a speaking personality and an oxygenated, breathing body. The extent to which we are willing to go beyond a reality principle in which only humans can be speaking personas depends on our willingness to refract ourselves within another logic. Superficially, this is a logic of commerce: 'What do people want from a virtual assistant?' asks Liz Stinson in a critique of such innovations' hidden commercial drivers:

Amazon's efforts to make Alexa sound as human as possible suggest that users expect their artificially intelligent sidekicks to do more than turn on their lights or provide a weather forecast. They want these devices to understand them. Connect with them. Maybe even – don't laugh – date them . . . Amazon wants to sell you things . . . and a more emotive assistant could be leveraged to that end . . . [I]t stands to reason that an AI more capable of expressing emotions would also be more capable of analyzing – and manipulating – your own. Creepy, yes, but also promising . . . Amazon might use Alexa's expressiveness to sell you stuff, but social robots could use the same technology to deliver, say, better care to the elderly.<sup>8</sup>

At another level there is a logic of innovation. Speech synthesis typically functions by sampling large amounts of recorded speech fragments from one individual so words can be reassembled into an utterance appropriate to the message being conveyed. Singing has proved equally susceptible to this kind of synthetic generation. Yamaha's Vocaloid software (2000–), fronted by teenage avatars such as Hatsune Miku and Megurine Luka, allows users to enter melodies and lyrics that are then 'sung' by a voice generated from a bank of samples. Recent commentators have sought to hear such sounds on their own terms as 'a real synthesized voice' rather than a stand-in for an actual human voice, an authenticating gesture that locates the software itself at 'the intersection between technology and creative production' (Jackson and Dines 2016, 108). While these remain rooted in human sounds, cobbled together by algorithm, synthetic voices emanate more fundamentally from the imitation of a human voice generated from raw waveforms. At present, the only example is Google's WaveNet wherein – like Markov chain melodies – a predictive distribution for each audio sample is conditioned on all previous ones, rising to at least 16,000 samples per second, all in pursuit of 'subjective naturalness'. This artificial approach to natural voices 'directly model[s] the raw waveform of the audio signal, one sample at a time' – we learn – 'As well as yielding more natural-sounding speech, using raw waveforms means that WaveNet can model any kind of audio,

including music.<sup>9</sup> Six five-second samples of intriguing, Skryabin-like piano music were available online at the time of writing; their brevity suggests style imitation via sampled recordings is achievable through automatic splicing, but perhaps little else as yet.

Why does it matter how human-like these machinic voices can sound? One answer concerns the reflexive identity such sound technologies afford us, where digital voices become a metonym for human–device relations more broadly. Sifting data on the Internet has allowed search engines to accrue sufficient algorithmic sophistication for users to treat Google’s search box as pilgrims once treated the high priestess Pythia, the fabled oracle of Apollo’s Temple at Delphi: a venue for the self’s unknowability, an intelligence advising individuals with seemingly better insight into them than they themselves possess. Rather than seeking prophesy, twenty-first-century users seek moral as well as informational guidance. In recent years popular questions to Google reported in the *Guardian* newspaper range from the personal (am I normal / why don’t you love me / why don’t people like me / am I a bad mother / why don’t I enjoy life) to the fantastical (why don’t unicorns exist) and the wryly speculative (are blond men evil).<sup>10</sup>

One reading of this practice is that digital search has encroached on the role of conscience, that ‘inner voice’, if only to cross-reference ours with those of others. Such a voice has at least four attributes: it is an inner guide and principle, it is acousmatic and not acoustically natural, it answers questions not immediately within the reach of our conscious reason, and –in most cases –it responds when consulted. It takes a moment to accept that all four of these attributes can be ascribed to the voices of the digitally synthesised AI bots above. Likewise, instantaneously cross-referencing such online ‘voices’ through search engines mimics the process of Google’s speech synthesis at 16,000 samples per second. Yet within a humanist tradition, *inner* voice is nothing less than ‘the link with God’, an ethical force genetically related to the Socratic voice, the ‘voice of the daemon that accompanies Socrates throughout his life’ (Dolar 2006, 83, 86). Consider the case of Jean-Jacques Rousseau, for whom an interior voice of conscience is a *moral* voice, and as such the marker of a common humanity:

Conscience! Conscience! Divine instinct, immortal and celestial voice; firm guide of an ignorant and limited being, but one which is also intelligent and free; the infallible judge of good and evil, it is you that make man similar to God . . . without you I do not sense anything in myself which would elevate me above the beasts, just the sad privilege to stray from error to error with the help of an intelligence without a rule and a reason without a principle. (Rousseau 1969, 600–1)

This vaunted status for inner self-reflection perhaps explains why, historically, simulating a human voice has been a goal of technologists ever since the sound of the voice was linked to the soul in its morphological likeness to breath (the Stoic's *pneuma*). La Mettrie's excited prediction in 1748 that 'a speaking machine . . . can no longer be considered impossible, particularly at the hands of [Jacques de Vaucanson], the new Prometheus' seemed at the time a final test of his over-arching proposition that people are inherently mechanical, but key to this reading is the significance ascribed to breath and his argument that humans have no soul as such (Mettrie 1996, 34). That is, he wasn't really subsuming the human within the mechanical, but vice versa.

Without necessarily buying into la Mettrie's materialism, recent digital simulations of voices have elicited both musical and data-driven responses. Within the context of popular musics, distortions of speech have formed a trope of the 'machinic voice'. This reverses the machine mimicry of human speech given above, for it distorts human sounds to mimic the cultural persona of the robot. A leading example is the vocoder (a signal-processing algorithm, formerly a military technology to mask speech communications during the Second World War (Tompkins 2011)); this is recurrent in pop recordings, from Neil Young ('Transformer Man', 1982) to Kanye West ('Love Lockdown', 2008), and can even be seen in blockbuster films, e.g. the glittery silver Guy Diamond in Disney's *Trolls* (2016), spoken by Kunal Nayyar entirely through vocoder distortions. Back in 2003, Joseph Auner speculated on how vocoder and computer simulations of voices play on the associations of mechanical and organic sounds in songs by Radiohead and Moby. Far from deconstructing the human, these present songs as 'a sort of cyborg system that attempts to splice the human and technological thus . . . illuminat[ing] its peculiar expressive character'. In other words, 'human' remains the dominant sign, against which 'cyborg' tensions and is defined. The ensuing anxiety of identity is embedded in the manipulation of vocal signifiers within a continuum of human and synthetic computer sounds. For Auner, the resulting cyborg persona 'becomes a way of reconstructing expression', which is to say, both a topos of pop culture and a referential language (Auner 2003, 110–11).

Within an art music tradition, Harrison Birtwistle's opera *The Mask of Orpheus* (1986) pioneered the use of an electronic voice for Apollo, 'speaking' in an invented language, whose otherworldly utterances were coded by Barry Anderson at IRCAM to sound god-like. By contrast, composer Peter Ablinger's 'Deus Cantando (God, Singing)' (2009) is only one of the most recent spectral analyses of recorded speech that forms the basis of his 'speaking piano', a computer-controlled player piano that



replicates on the instrument's eighty-eight keys the decomposed sound spectrum of recorded human speech:

Using . . . 16 units per second (about the limit of the player piano), the original [sound] source approaches the border of recognition within the reproduction. With practice listening[,] the player piano can even perform structures possible for a listener to transpose into/understand as spoken sentences.<sup>11</sup>

That is, you can 'hear' the piano pronounce words only when you simultaneously see its words or know them in advance. This accommodation of digital sound file and keyboard mechanism has been enlisted to present analogue voices of the dead – Schoenberg, Brecht – 'speaking' in the present, and, to that end, replicates one of the initial functions of Edison's crank-driven phonograph, but remediated through spectral recomposition of the source: a form of near-human expression in which the digital (electricity) accommodates the mechanical.

In an experiential sense, such sounds perform a kind of time travel: the piano's hammers 'speak' Schoenberg's words mechanically in the present. And here it is worth reminding ourselves that mechanical work – from instrument building to repetition in the practice room – is predicated on a principle of fragmentation whereby larger tasks are divided up into smaller tasks that can be performed in linear sequence. As David Hulme first noted, however, there is no principle of causality in a mechanical sequence; movement through the sequence generates change without oversight of how operations follow one another. It was this blindness within mechanisms that led McLuhan to accord electricity the crown of all industrial and post-industrial inventions, for 'it ended sequence by making things instant' (like Brecht and Schoenberg's spectrally recomposed voices in the here and now). Hence his cryptic assertion that the electric light 'is pure information' (McLuhan 1964, 8, 12). If the electric grid underpinned McLuhan's 'global village' of instant communications, it also models the modern network, so, beyond the supply of power, electricity itself remains regulative of an aesthetics of digital media: 'deeply shaping of the form and content of the medium itself' (Dewdney and Ride 2006, 79). Record companies regularly take advantage of this time-travelling instantaneity in digital editing by reusing the recordings of Pavarotti, Sinatra, Nat King Cole, Tupac Shakur and others and ventriloquising new duets or new backing after death, in a temporal short-circuit that appears to make these voices sing anew. By comparison with the macro-structure of the Internet, Ablinger's speaking piano is philosophically significant, in part, because of its equal reliance on data networking *and* old-fashioned mechanical keys, hammers and physical resonance. From this arises a paradox: just as the concept of the posthuman is ultimately embedded in and defined against the human, so the virtual is here embedded in the material.

### Case Study: *Death and the Powers* (2011)

But this apparently comfortable accommodation between the human and the posthuman turns out to be unsustainable in the end, a point aptly demonstrated by a recent piece of music theatre. Beyond the semiosis of robotic voices, we now turn to what might be called a ‘theatrical’ semiosis of human–digital music relations. Tod Machover and Robert Pinsky’s *Death and the Powers: A Robot Pageant* is a one-act opera ‘conceived and written specifically for the incorporation of new technologies . . . that re-envision human presence onstage’, explains Peter Torpey, a student designer of its multimedia systems (Torpey 2012, 110). It stages the tensions and imaginative possibilities of a posthuman discourse vis-à-vis digitised music, reflecting on mortality as alibi to technological necessity. As a narrative, it depicts a critical juncture in the vision of our postbiological future first put forward by Hans Moravec: the downloading of our consciousness into an all-powerful computer system, and the discarding of our original, mortal body, leading to the extinction of the human race (Moravec 1988, 112). Tellingly, the protagonist becomes nothing but a digitised, acousmatic voice, though we are to understand that he retains the legal, fiscal and moral authority he held as a bodied human. The opera emerged via MIT media lab’s ‘opera of the future’ project in 2010, and has been performed in Monaco, Boston and Chicago to date.

First, a brief plot summary. The narrative is told from the distant future by robots who enact the ‘pageant’ in order to try to understand what biological death is, and how the human race died out. At the outset, four robots debate the mystery like latter-day paleontologists:

What is this ‘Death’ – Is it a form of waste? / . . . I cannot understand. / If the information of one unit might be lost / it is backed up by any other unit at hand.  
 Death – Is it an excessive cost? . . . Is it the data rearranged, / As in an error, in a dream? . . . A dream of something lost / That was meant to be saved? (ll. 8–32)

The robots then transform themselves into the ‘human’ characters on stage, and the action commences: Simon Powers, an aged, wealthy, wheelchair-bound inventor and magnate, is planning his death as a transmutation of himself into the ‘System’. He explains his philosophy of embodiment to worried relatives by emphasising that ‘[i]t’s the vibration, / The movement, that matters! / . . . It’s never matter that matters. Particles, molecules, cells, fingers, eyes, nerves / Are only places for the system / Of meaningful vibration’ (ll.122–47). Simon’s third wife (Evvy) and daughter (Miranda) are sceptical (‘But how can you be yourself / Without a body?’ ll. 212–13), but receive only cryptic answers in the form of poetic quotation.<sup>12</sup> In the second scene, Simon’s body disappears as he utters an aphoristic existential valediction: ‘remember: / Whatever I did / I *did* that and /

I am the same' (ll. 279–81). While waiting for Simon to emerge from the System, his adopted son and protégé, Nicholas, reflects on his own prosthetic enhancements and how he helped his father to live inside the system. Simon soon emerges from a portrait of his younger self (initially as a hologram, later as pure voice), moves around mechanically, and eventually asserts his new identity:

What is my name? / A name is a machine. / A name is a made-up thing /  
That proposes someone is real. / My name is a machine for designation –  
That's what any name is. / My name is Simon Walter Powers, / It proposes  
I am alive. (ll. 347–55)

The location of his voice shifts within the scene (using 140 speakers and spatial diffusion – Wave Field Synthesis – to pinpoint sound on stage), between a portrait, a mechanical bird perched on set, and elsewhere in the room. His relatives debate whether this disembodied entity is still 'Simon'. Evvy later seeks intimate contact (and thereafter is reduced to wordless utterances), and the family is visited by a delegation from the 'United Way, / the Administration / And the United Nations' (ll. 476–8), who seek help to combat crises of world famine, biological weapons, child exploitation and climate change in the wake of Simon's sudden financial liquidation. Simon merely quotes 'O Röschen rot' from *Des Knaben Wunderhorn*, reinterpreting the lines set by Mahler without melodic quotation: 'I am from God and will return to God, / Dear God will give me light, / Will light me to eternal life!' (ll. 602–4).<sup>13</sup> The delegation's stupefied response is met with an anecdote about meaningless violence among today's youth, and they suspect Simon may be dead, his voice a trick. Bending Mahler's symphonic resurrection to the ecumenical present, Simon misreads the original poetic line by replacing 'God' with 'light' – I am from light and will return to light – as his voice is frequency modulated to extend into an artificially high register in a modified rising whole tone scale. Before entering the System, Evvy declares the feeling inside it as a giddy sense of unending freefall, perhaps gesturing to the 'bodiless exaltation of cyberspace' (Gibson 1984, 123). Nicholas, who has been increasing his prosthetic enhancements, follows suit, leaving Miranda alone in her human skin. She reflects on the unethical escape 'into the light' (l. 771) of the few, and the pain of the millions, before Simon returns as a hologram in a wheelchair (a wry technology joke) to explain the rationale for evolving into non-biological forms:

Like you, I tried to help the world. / I, too, saw these miseries . . . / But the  
animal is defective. / . . . We evolved as meat, to love fat and sugar; / Once  
that was good, but now it is fatal. / We evolved as flesh to want sex all the  
time; / Once that was good, but now it is fatal. / We evolved as muscle to

want to make war; / Once that was good, but now that is lethal . . . Now  
there's no help but evolving / Out of the meat, and into the system. / It isn't  
the many and the few – / It's yourself, it's you! (ll. 780–807)

Miranda's protest that misery is part of human identity – a last stand of humanism – is followed by her anxious contemplation of a future alone, without a huggable mother or father, with 'no lover, no other' (l. 822). She defends the body, death, sugar and meat, closing with a battery of rhetorical questions.

In the epilogue, the robot-actors return to character as 'operabots' and repeat the questions of the prologue, still with no answers. The lead robot reiterates a message whose mindless reiteration acquires a sinister ring, given the robots' final incomprehension of human empathy: 'Units deployed as Individuals will receive / One Thousand Human Rights Status Credits' (ll. 924–5).<sup>14</sup> The monotonous message, delivered 'dry, no emotion, no vibrato', imitates the cultural topos of the robot, confirming that all prior expression had been unreal, a calculated simulation of human expression.

But the work performs its posthuman identity in two senses: by inviting us to empathise with robot actors ('operabots') playing humans, we already reach across the alterity relation within the opera's narrative, even though the robot characters are in fact played by human singers. The double impersonation (human impersonating AI device impersonating human) complicates the usual means of differentiating AI from human (the so-called Turing test), for the established circularity is theoretically endless: an identity multiplied *ad absurdum*. At what point, in other words, does impersonation end and identity begin? Such ambiguity would seem precisely the point in the impulse to render virtual speaking assistants increasingly human-like. On the one hand, we know when the principal singer, Powers, is offstage, it is the human actor's movements and breathing – detected in a booth by wireless sensors and filtered through algorithms – that determine the vocal amplification and stage environment: its lighting, movement of props, stage scenery, etc.

Data from these sensors and the singer's voice are streamed to custom software for analysis and then used to drive and influence motion, illumination and visuals throughout the theatrical environment onstage that accompanies his amplified singing voice. (Torpey 2012, 115)

Hence the agency driving the stage effects lacks any intentionality in performance (physical gestures offstage translate into onstage visual effects, but not in a way that the singers can control), and to that end might be considered more a distributed cognition than a human performance. On the other hand, at a different level of realism this is nothing but

concealment, for humans must first learn the score and programme the algorithms governing movement-driven stage effects. In this sense, the doubly suspended disbelief required of an audience receives its complement in the contradictory stage identities that remain suspended between robotic and human singers.

Staging robotic voices in this way, measured against an index of human likeness, dramatises the relation between human and digital technology as an agon, a conflict that has no end in sight. This dark ontology of digital technology has lurked on the periphery of science fiction for decades, yet its menacing predictions of loss and alienation appear no closer to fulfilment. For sound technology the reciprocal paradigms of mimesis (composing with algorithms) and semiosis (synthetic voices) explored in this chapter present two ways of relating AI to human identity. Accepting relations with devices is a fact of digital ‘assistance’ that we cannot do without; how the indices for determining and evaluating these relations are chosen remains a matter of debate, a debate whose framing parameters are unclear.

Media devices appear to offer a veil of neutrality, for they make no distinction between or judgment on the sound sources they engage: animal or human, naturally occurring or artificially produced, pop music, art music or military explosion. There is only frequency response, bandwidth and transistor processing speed. Yet devices have affordances that shape the experience of users. So the flipside is that increasingly we are unaware of the digital hand guiding our musical experience.

As illustrated by the sound and staging of digital voices, empathy generation is at present a heavily gendered, commercial enterprise, from Hilary’s respect for cognitive psychology in Stoppard’s play, to the ubiquity of female-sounding digital assistants, and the opposition of a female-body versus male-brain (Miranda, Simon) in Machover’s opera. The extent to which we are troubled or indifferent to this matters less, perhaps, than the knowledge that the role of digital media in musical creativity can only grow, and with it the responsibility to monitor such developments.

### **For Further Study**

- Auner, Joseph. 2003. ‘Sing it for Me: Posthuman Ventriloquism in Recent Popular Music’. *Journal of the Royal Music Association* 128 (1): 98–122.
- Bostrom, Nick. 2009. ‘The transhumanist FAQ’. In *Readings in the Philosophy of Technology*, edited by David Kaplan, 345–60. 2nd edn. Plymouth: Rowman & Littlefield.
- Hayles, Katherine. 2012. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: Chicago University Press.
- Nierhaus, Gerhard. 2009. *Algorithmic Composition: Paradigms of Automated Music Generation*. Vienna: Springer.

- Stoppard, Tom. 2015. *The Hard Problem*. London: Faber.
- Torpey, Peter A. 2012. 'Digital Systems for Live Multimodal Performance in *Death and the Powers*'. *International Journal of Performance Arts and Digital Media* 8 (1): 109–23.

## Notes

- 1 Elle Hunt, 'Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter', *The Guardian*, 24 March 2016, [www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter](http://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter). All websites accessed 7 September 2018.
- 2 Madhumita Murgia, 'Microsoft's racist bot shows we must teach AI to play nice and police themselves', *The Telegraph*, 29 March 2016, [www.telegraph.co.uk/technology/2016/03/25/we-must-teach-ai-machines-to-play-nice-and-police-themselves/](http://www.telegraph.co.uk/technology/2016/03/25/we-must-teach-ai-machines-to-play-nice-and-police-themselves/).
- 3 The shift 'from corporeal communication to disembodied data' that is commonly associated with posthumanism is also discussed by Paul Sanden (Chapter 7, this volume, pp. 184–5); for Sanden, 'liveness functions as an index of humanness'.
- 4 Dmitri Kartofelev and Jüri Engelbrecht, 'Algorithmic melody composition based on fractal geometry of music', presentation, August 2013, [www.cs.ioc.ee/~dima/fractalmusic.html](http://www.cs.ioc.ee/~dima/fractalmusic.html).
- 5 Jefferson Graham, 'Someday, Amazon wants you to have long talks with Alexa', *USA Today*, 2 May 2017, [www.usatoday.com/story/tech/talkingtech/2017/05/02/whispering-alexa-just-start-says-amazon-head-scientist/101171340/](http://www.usatoday.com/story/tech/talkingtech/2017/05/02/whispering-alexa-just-start-says-amazon-head-scientist/101171340/).
- 6 Brent Rose, 'Stand-up comedy using only Siri, Alexa, Cortana and Google Home', *Wired*, [www.youtube.com/watch?v=rO-89oBeBbQ](http://www.youtube.com/watch?v=rO-89oBeBbQ).
- 7 Freia Lisa Lobo, 'Alexa can whisper now', *SoundCloud*, [soundcloud.com/freia-lisa-lobo](http://soundcloud.com/freia-lisa-lobo).
- 8 Liz Stinson, 'The surprising repercussions of making AI assistants sound human', *Wired*, 5 May 2017, [www.wired.com/2017/05/surprising-repercussions-making-ai-assistants-sound-human/](http://www.wired.com/2017/05/surprising-repercussions-making-ai-assistants-sound-human/).
- 9 Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Viuyals, Alex Graves, Nal Kalchbrenner, Andrew Senior and Koray Kavukcuoglu, 'WaveNet: A generative model for raw audio', 19 September 2016, [deepmind.com/blog/wavenet-generative-model-raw-audio/](http://deepmind.com/blog/wavenet-generative-model-raw-audio/).
- 10 Various, 'The autocomplete questions', *The Guardian*, [www.theguardian.com/commentisfree/series/the-autocomplete-questions](http://www.theguardian.com/commentisfree/series/the-autocomplete-questions).
- 11 Peter Ablinger, 'Quadraturen', <http://ablinger.mur.at/docu11.html#principles>.
- 12 The two poems Simon uses are Yeats's 'Sailing to Byzantium' (1924) and May Swenson's 'Question' (1954).
- 13 See 01:08–02:23 in 'Tod Machover's *Death and the Powers* Scene 6 – The World Pleads', [www.youtube.com/watch?v=w3zC7bf7qsU#t=1m8s](http://www.youtube.com/watch?v=w3zC7bf7qsU#t=1m8s).
- 14 See 01:40–01:50 in 'Tod Machover's *Death and the Powers* – Prologue', [www.youtube.com/watch?v=Myd2DdSxUEk&list=RDMyd2DdSxUEk#t=1m40s](http://www.youtube.com/watch?v=Myd2DdSxUEk&list=RDMyd2DdSxUEk#t=1m40s).