

CLASSICAL MUSIC FUTURES PRACTICES OF INNOVATION

EDITED BY
NEIL T. SMITH, PETER PETERS,
AND KAROLY MOLINA





<https://www.openbookpublishers.com>

© 2024 Neil Thomas Smith, Peter Peters and Karoly Molina



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (CC BY-NC 4.0). This license allows you to share, copy, distribute and transmit the work for non-commercial purposes, providing attribution is made to the author (but not in any way that suggests that he endorses you or your use of the work). Attribution should include the following information:

Neil Thomas Smith, Peter Peters and Karoly Molina, *Classical Music Futures: Practices of Innovation*. Cambridge: Open Book Publishers, 2024,
<https://doi.org/10.11647/OBP.0353>

Further details about CC BY-NC licenses are available at
<https://creativecommons.org/licenses/by-nc/4.0/>.

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at
<https://archive.org/web>.

Copyright and permissions for the reuse of many of the images included in this publication differ from the above. This information is provided in the captions. Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notification is made to the publisher.

Any digital material and resources associated with this volume may be available at
<https://doi.org/10.11647/OBP.0353#resources>.

ISBN Paperback: 978-1-80511-073-6

ISBN Hardback: 978-1-80511-074-3

ISBN Digital (PDF): 978-1-80511-075-0

ISBN Digital eBook (EPUB): 978-1-80511-076-7

ISBN XML: 978-1-80511-078-1

ISBN HTML: 978-1-80511-079-8

DOI: 10.11647/OBP.0353

Cover image: Jean Pierre Geusens, *Focuss22* (2017)

Cover design by Jeevanjot Kaur Nagpal

19. Ghosts of the Hidden Layer

Jennifer Walshe

The following was originally delivered as a talk at the Darmstädter Ferienkurse in 2018.

Today I'm going to talk about the voice, language and artificial intelligence. We're going to cover a lot of ground, but at the end we should all end up in the same place.

1.



Abstract sound, mixture of static and wind-like sounds. Duration: 0:16.
<https://hdl.handle.net/20.500.12434/6227db0e>



1st Audio: Dadabots © Jennifer Walshe.

I don't have a voice. I have many, many voices. My voice – this biological structure located in my body, an apparatus which usually functions in close collaboration with technology – is the staging area for everything I've ever heard and everywhere I've ever lived. There is infinite material to draw on, in infinitely different ways.

We all grapple with the plethora of voices that have made their mark on ours. We're told the goal is to find "our" voice. But this polyphony, this confusion, is what interests me. I don't want to choose.

I listen to and collect voices constantly. Recording, notating, jotting down times as I watch a video so I can go back and memorise voices.

A lot of my work deals with negotiating and editing this huge archive of material. The first piece I ever wrote for my voice, as *mo cheann*, was

concerned with investigating all the vocal sounds I could make along a vertical physical axis, projecting from the space above my head to below my feet. I've made pieces for my voice which represent the connections between the thousands of pop song samples in my brain; I've made pieces which are the result of hours upon hours of listening to animal, insect and frog sounds; pieces for which recordings of DNA microarray machines, underwater volcanoes and the toilets on the International Space Station form an aural score.

Because I love voices, I love language. I view language as a subset of what a voice does. I am fascinated by how language functions off-and online. I love slang and argot technical language; I love newly invented words. Through language, voices give a vivid snapshot of the times we live in. Times filled with collarbone strobing, meta predators and procrastibaking. Hate-watching, nootropics and dumpster fires. Manbabies, co-sleepers and e-liquids. I read these words on the page and they bounce into life in my head as voices.

Aside from the extremely rich sound world the voice can be implicated in, I'm fascinated by the voice because it provides an aperture through which the world comes rushing in. The Thai artist Larry Achiampong states that 'Our lives are political because our bodies are.' I would extend this statement and say our lives are political because our voices are. Gender, sexuality, ability, class, ethnicity, nationality – we read them all in the voice. The voice is a node where culture, politics, history and technology can be unpacked.

2.



Peaceful melody with choral accompaniment. Duration: 0:37.
<https://hdl.handle.net/20.500.12434/342be800>



2nd Audio: EMERALDIA 'Llac Centre Suite' (excerpt) – Celtorwave track
 © Jennifer Walshe.



Fig. 19.1 *Celtorwave* by Tom Kemp @ Jennifer Walshe.

In January 2015 I launched a project called Historical Documents of the Irish Avant-Garde, also known as Aisteach. For this project, I worked in collaboration with a wide number of people to create a fictional history of Irish avant-garde music, stretching from the mid-1830s to 1985. Many of the materials related to this project are housed at aisteach.org, the website of the Aisteach Foundation, a fictional organisation which is positioned as 'the avant-garde archive of Ireland.' The site contains hours of music, numerous articles, scores, documents and historical ephemera. Every detail of this project was composed, written and

designed with the utmost care and attention to detail. It's a serious exercise in speculative composition, fiction and world-building. In Aisteach, Irishness becomes a medium. Aisteach creates an uncanny space, where we write our ancestors into being in the hopes of being able to summon their voices and listen to them.

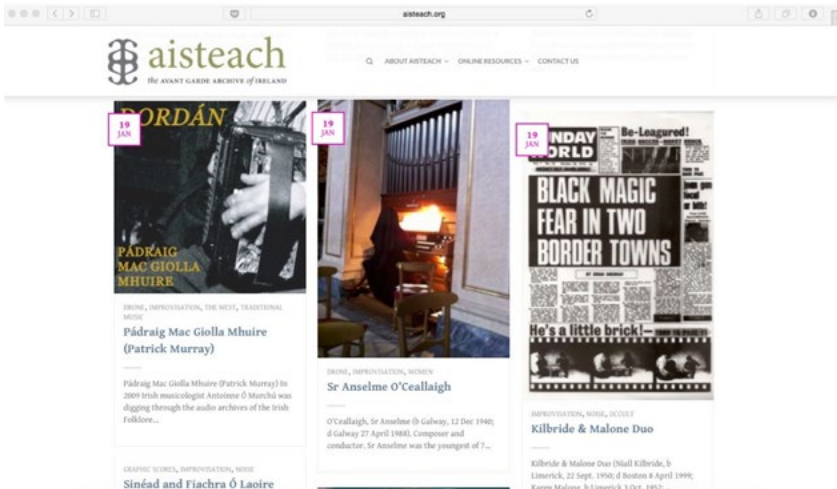


Fig. 19.2 Aisteach © Jennifer Walshe.

Aisteach includes many different musics, many personae¹. The subliminal tapes and films of Outsider artist Caoimhín Breathnach, my dear great-uncle; The Guinness Dadaists, staging sound poetry performances in the brewery at St James's Gate; Sr Anselme, a nun living in an enclosed order, immersed in long-form organ improvisations; Roisín Madigan O'Reilly, the radio enthusiast performing with ionosphere and landscape; Dunne's Dérives, the notorious queer performance nights featuring members of the Radical Faerie movement.

Aisteach is an on-going project. It continually spills over into the offline world in the form of concerts, film screenings, exhibitions, radio shows.

Two of Aisteach's most respected fellow travellers include Afrofuturism and Hauntology. Hauntology in particular seems like

¹ For further reading about the musics and personae, see aisteach.org.

an obvious native choice for Irish-based art, embedded as we are in superstition, nostalgia and the occult. But Hauntology functions differently in Ireland – Aisteach is haunted by a past which suppressed, marginalised and erased many voices. Aisteach is not interested in fetishising this past. The crackle on the recordings is not there for cosy retro warmth or nostalgia for the rare oul’ times – it’s sand on the lens, grit between the tape heads, violently hacking history to urge us to create a better future. And a better future means being alert and responsible to the present.

At the moment I’m preparing a new Aisteach exhibition.² We’re working on Celtorwave, an Irish version of the online subculture Vaporwave; we’re collecting water from holy wells for a holy water cocktail bar; we’re building an artificial intelligence system that writes Irish myths.

3.



Generated sound with Walshe’s voice of static-like whistling sound.
Duration: 0:18.
<https://hdl.handle.net/20.500.12434/fb0d7fd8>



3rd Audio: Jennifer Walshe’s voice generated by Dadabots, Epoch 6
© Jennifer Walshe.

On September 11th 2001, the American-Canadian science fiction author William Gibson was at home, drinking coffee, when he heard of the attacks on the World Trade Center. He describes how he ‘ran upstairs, turned on CNN and that was it. “All bets are off!” my inner science-fiction radar screamed. “Cannot compute from available data.”’

When the attacks happened, Gibson was one hundred pages into writing a new novel. Weeks after the attacks, as Gibson tried to get back to work, he realized his work in progress ‘had become a story that took place in an alternate time track, in which September 11th hadn’t happened.’ He had to rework the novel completely.

2 Jennifer Walshe, *The Journal of Music* (2023), <https://journalofmusic.com/subject/jennifer-walshe>

Gibson is one of my favourite authors, and in June 2016 and again in November I wondered what he was doing. How many pages he had to rip up. How many plotlines he had to rework.

This is what I'm interested in – making work which is concerned with the world that we're living in. Work which thinks deeply about and through the human and non-human beings in this world and the universe beyond it. Work that is so enmeshed with and affected by the world that it cannot help but change in response to it.

Gibson says the approach to doing this is to use the sci-fi toolkit, to use 'science fiction oven mitts to handle the hot casserole' of the times we live in.³ Because our world is utterly strange, utterly more bizarre, layered and textured than any science fiction scenario or imaginary future could be. And a huge part of that strangeness is because we are overwhelmed, dominated and enmeshed with technology.

4.



Walshe's synthesised voice saying, "If I was dying and I had a turtleneck on, he tells me, with my last dying breath, I would take the turtleneck off and try to throw it as far away from my body as possible." Duration: 0:09.
<https://hdl.handle.net/20.500.12434/e913fb49>



4th Audio: Elon Musk quote spoken by version 1 of Walshe's synthesised voice
 © Jennifer Walshe.

The concept of the 'uncanny valley' was first described by the Japanese robotics professor Masahiro Mori in the journal *Energy* in 1970.⁴ Mori's paper deals with the challenges facing designers of prosthetic limbs and robots. Mori describes a graph with two axes – affinity and human likeness. Mori's model describes how we feel little affinity with industrial robots which in no way resemble people, versus feeling huge affinity with healthy humans.

3 Douglas Gorney, 'William Gibson and the Future of the Future' in *The Atlantic* (2010), <https://www.theatlantic.com/entertainment/archive/2010/09/william-gibson-and-the-future-of-the-future/62863/>

4 Masahiro Mori, 'The Uncanny Valley: The Original Essay by Masahiro Mori' in *Spectrum* (2012) <https://spectrum.ieee.org/the-uncanny-valley>

Between these two extremes, we find the uncanny valley. As robots, or any other representations of humans, such as dolls or puppets, come to more closely resemble humans, we feel increasing levels of affinity, until we come to the uncanny valley, at which point humans become completely freaked out. The uncanny valley is inhabited by corpses, zombies, prosthetic limbs and robots who are 'almost human' or 'barely human.' For Mori, as we enter the uncanny valley there is an abrupt shift from empathy to revulsion.

Mori theorises that the sense of eeriness we feel in the uncanny valley is without a doubt 'an integral part of our instinct for self-preservation.' Mori encourages the building of 'an accurate map of the uncanny valley, so through robotics research we can come to understand what makes us human.' His goal is a compassionate one – by understanding the uncanny valley, designers can make better prosthetic limbs; prosthetics which will put their users and the people around them at ease.

Japanese robotics professor Hiroshi Ishiguro designed the android Geminoid to look exactly like him, going so far as to implant some of his own hair in the Geminoid's head.

Developing realistic-sounding voices and coupling these with realistic-looking movements is one of the greatest challenges facing robotics. The Geminoid has a speaker in its chest. Ishiguro speaks into a microphone, and his voice is relayed through the speaker as the Geminoid approximates the mouth movements implied by the speech. This is where the illusion starts to fade most rapidly.

The Otonaroid, another of Ishiguro's robots, is on display at the Museum of Emerging Technology in Tokyo. When I visit, I speak to her, and her voice comes out of a speaker on the wall behind her as she gestures independently. It is confusing and disorienting. When I pet Paro, the robotic therapy seal, and it purrs back to me, I can feel the parts of its body where its speaker is located. It is both the same and also entirely different to petting a live, purring cat.

5.

When I was eleven, my father worked for IBM.⁵ He took my sister and I to an event for the children of IBM employees. This event was extremely exciting, because IBM had hired a Hollywood actor to come and provide entertainment for the kids. The actor IBM hired was the actor Michael Winslow, known as 'The Man of 10,000 Sound Effects'. Winslow is famous for his stunning ability to mimic sounds with his voice. He played the character of Larvell Jones in all seven of the Police Academy movies.⁶ I sat spellbound as Winslow used his VOICE to make the sounds of telephones, engines, sirens and tape recorders. He was amazing. I witnessed a human using their voice to disrupt, confuse and explode notions of embodiment, of what the voice and the body are.

Winslow was a master of extended vocal techniques. But there was one sound he could not do. Winslow could not do an Irish accent. We witnessed him fail that day. But he was no less a hero to me.

Irish people have grown used to seeing people crucify the Irish accent. Tom Cruise in *Far & Away*.⁷ Ryan O'Neal in *Barry Lyndon*. There is a circle in hell, a neon green, shamrock-encrusted Irish pub in Sunnyside, where Seán Connery is condemned to eternally do dialogue from *Darby O'Gill & the Little People*.⁸ But not only did we grow up watching people failing to do Irish accents in television programs and films; we also saw it happening on the news.

Between 1988 and 1994 the British government under Margaret Thatcher banned the broadcast of the voices of members of Sinn Féin and other Irish republican and loyalist groups. This meant that broadcasters could show footage of someone such as Gerry Adams speaking, but they

5 Céad Míle Fáilte – Welcome to Ed Walshe's Kayak Site. <http://homepage.eircom.net/~sizler/index.html>

6 Movieclips, *Police Academy (1984) – Larvell Jones, M.D. Scene (1/9)*, movie clip, YouTube, June 19, 2019, https://youtu.be/6OKt2CZ4ULE?si=D-KBfCV9gL5xb9_g

7 Movieclips, *Far and Away (4/0) Movie CLIP – Say You Like My Hat!*, movie clip, YouTube, June 16, 2011, <https://youtu.be/eKrEVWGTuRg?si=PDHmHoDHpjLjVEmu>

8 Bmoviereviews *Darby O'Gill and the Little People – Sean Connery's Leprechaun Dream*, movie clip, YouTube, June 30, 2010, <https://youtu.be/--DYI9wTtAs?si=Chbl3nRSBY5QfANg>

could not broadcast the sound of his voice. Broadcasters got around the restrictions by hiring Irish actors to 're-voice' the original voices.

There were many approaches to the re-voicings – some actors over-acted in an attempt to get political points across; others attempted to be neutral; some journalists asked the actors to deliberately speak out of sync, to highlight the absurdity of the restriction.

Irish actor Stephen Rea, who was nominated for an Academy Award for his role in *The Crying Game*, re-voiced both Gerry Adams and Martin McGuinness. The process gave Rea powers that went way beyond traditional voice-acting – Rea has described how he tried to make Adams and McGuinness's messages as clear as possible by editing their speech during the re-voicing process, eliminating the hesitations, umms and aahs of the original. The irony is stunning – by choosing to literally silence the voices of republican and loyalist groups, the British government enabled a situation where world-class actors had the power to polish extremist voices and make them more eloquent.

What was happening, in cognitive terms, when we watched Martin McGuinness on TV? Whose voice was speaking when we saw his lips move? Was it Martin McGuinness? An actor deliberately speaking out of sync? Was it New & Improved Martin McGuinness, courtesy of Stephen Rea? Was it Margaret Thatcher, the prime minister who introduced the ban? The uncanny valley explodes into the political realm.

6.

The virtual digital assistant market is projected to be worth \$15.8 billion by 2021. As voice interaction is central to virtual digital assistants, all of the major tech companies are currently investing huge sums in voice technology – Amazon alone have created a \$100 million Alexa fund. The thirty-six percent of the world's population who own smartphones have access to virtual assistants such as Siri, Google Assistant, Cortana, Samsung S Voice.

Voice assistants like Siri or Cortana use concatenative text to speech – they sew together fragments from pre-existing recordings of human speech. Concatenative text to speech relies on huge databases – each individual voice is the result of one person spending days recording

thousands of words. It sounds somewhat natural, but has its limits – the database will not contain recordings of every word in current use, and switching to a new voice means recording an entirely new database.

The holy grail of voice synthesis is natural-sounding speech that does not require huge databases of recordings. Voices that can be expressed as models in code, models that can be modified with infinite flexibility.

In 2017, a group of researchers from MILA, the machine-learning lab at the University of Montreal in Canada, launched a company called Lyrebird (since absorbed by Descript).⁹ Lyrebird creates artificial voices. Using recordings of a person's voice, Lyrebird creates a 'vocal avatar' of that person. The recordings are not sampled – they are analysed by neural networks. Lyrebird's system learns to generate entirely new words – words that the original person may never have spoken in real life.

As soon as Lyrebird releases a beta version, I make a wide range of vocal avatars. Natural Jenny. Dubliner.J+. Jenpoint1000. I pump in text I've collected over the years. Ultimately, though, my feeling is of frustration. The rhythmic patterns of the voices are always the same. When I try to create a vocal avatar with radically different vocal cadence it crashes the system. I can hear a soft buzz whirring through every recording.

Let me be clear – Lyrebird is a huge technical accomplishment. But ultimately Lyrebird is not weird enough for me. I feel like I'm seeing brochures for the watered-down tourist section of the uncanny valley. I want the real thing.

Lyrebird named their company after the Australian lyrebird, a creature known for its stunning ability to mimic sounds. Lyrebirds have been heard making the sounds not only of other birds and animals, but also human voices, camera shutters and car engines. They do this extremely convincingly. Why do lyrebirds make these sounds? What do they think they're communicating when they reproduce the sound of a chainsaw? What do I think I'm communicating when I make animal sounds with my voice, which I do in so many of my pieces? I find the uncanny more readily in the biological lyrebird than the digital one. In

9 Masahiro Mori, 'The Uncanny Valley: The Original Essay by Masahiro Mori' in *Spectrum* (2012) <https://spectrum.ieee.org/the-uncanny-valley>

the tragedy of a wild creature imitating the sound of machinery that's cutting down trees in the forest it inhabits.

But this broader listening is where so much meaning resides. If the Broadcast Ban were to happen now, I could imagine the BBC hiring Lyrebird or Google to make a pristine model of Gerry Adams's voice. They would bypass the producers, the actors, all those messy humans who directly intervened in the re-voicing process. How would we listen to this voice? What would we hear? What would its politics be? And would we even hear the message, or simply be struck by the technological achievement? Would the most salient part of the voice be the cosy relationship between the state and the tech companies which dominate our lives?

7.



Fig. 19.3 Notebooks © Jennifer Walshe.

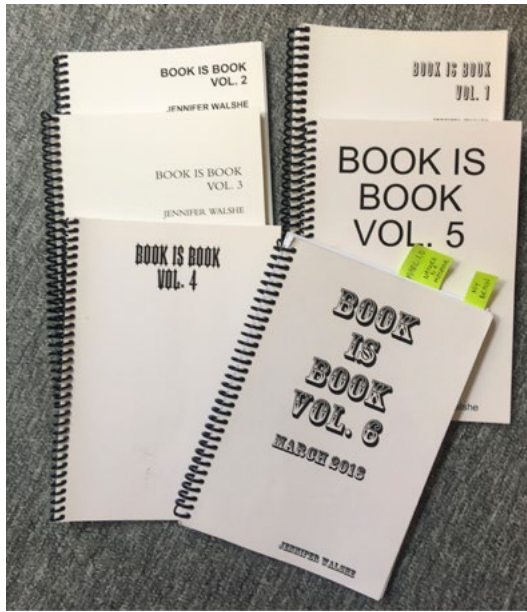


Fig 19.4 Book is book © Jennifer Walshe.

Since I was a child, I have collected text in notebooks. I didn't know of the tradition of commonplace books or about literary works like *The Pillow Book* when I started doing this. I just didn't want to lose anything – stories, poetry and songs I wrote, jokes my friends told, conversations I witnessed or simply overheard, lines from films or TV shows, text from books and magazines.

I started writing these notebooks by hand, but these days I collect everything on Evernote across all my devices. Every few months I edit the files, send them off to a print-on-demand service, then wait excitedly for the next volume to arrive in the post. I call this archive BOOK IS BOOK. When I'm composing, these books are close to hand. When I improvise using text, I use the books in the same way a DJ uses records.

Over the last few years, BOOK IS BOOK has been used as the input for various machine-learning projects. Bob Sturm and Oded Ben Tal, two composers and machine-learning specialists based in London, fed it into their neural network Folk-rnn.¹⁰ Folk-rnn was originally developed to write folk songs, but also work for text generation. The output is

¹⁰ folkRNN, <https://folkrrnn.org/>

exactly what I wanted – bizarre, with shades of Early English, *Finnegan's Wake* and Gertrude Stein:

Tumpet. Not be to strowded voice this singo to food so your before to days.
action and say enouginiens. To be the cingle from milodaticls get preference
to could, ever this experience 3 isfortation, melity, if I parky to before is
redelf winter, you've becomited specialised into a mecluate activatically

With output like this, my job is to commit. To sell it as if I understand it innately, and through that process gain a new understanding of what text and the voice can mean. The reality is that the second I read the text, it starts to infer, imply, even demand its vocal treatment. And that is a process driven by both myself and the neural network that produced it.

But this is the point – I'm interested in AI because I would like to experience not just artificial intelligence but also alien intelligence. Brand new artistic vocabularies, systems of logic and syntax, completely fresh structures and dramaturgies.

My role as an artist is to pay very close attention to the output of an AI, trying to understand and interpret this output as a document from the future, blueprints for a piece which I try to reverse-engineer in the present.



Sound poetry and vocal work by Jennifer Walshe. Duration: 0:38.
<https://hdl.handle.net/20.500.12434/437a7854>



5th Audio: THEY GO PEOPLE & PERSON (excerpt) by Jennifer Walshe.

8.

I love text scores, but I've found much of contemporary text score practice frustrating for a long time. In many scores, the vocabulary, the syntax, even the verb choices have remained static since the 1960s. For many of the composers of these scores, this simplicity is precise and apt. But I feel strongly that text scores are the most democratic, efficient, powerful form of notation, and yet we're stuck aping the linguistic style of the Fluxus period.

I feel we're missing out on a rich engagement if we fail to let text scores be affected by the language of the times we live in. I want text scores that are warped by Twitter and Flarf, bustling with words from

Merriam-Webster's Time Traveller, experimenting with language in the way Ben Marcus and Claudia Rankine do. If, as Donna Haraway says, 'Grammar is politics by other means', why not take the opportunity to interfere?

My early text scores were concerned with trying to push the limits of what a text score could do by turning it into highly technical language dealing with esoteric procedures. These experiments culminated in a large-scale score titled *The Observation of Hibernalian Law*. The work consists of a book, objects, diagrams and schematics.

In recent years I've been using the web to produce scores, trying to witness how the ecosystems of different social media platforms affects how a score is made and the sounds it might produce. I've made text-score projects on Snapchat, YikYak and numerous projects on Twitter.¹¹ These works aren't made for conceptual LOLs, they're serious attempts to welcome contemporary technology and language to the text score. The language in many of these scores was produced by feeding pre-existing text scores into Markov Chain Generators, by crossbreeding text scores and Weird Twitter accounts.

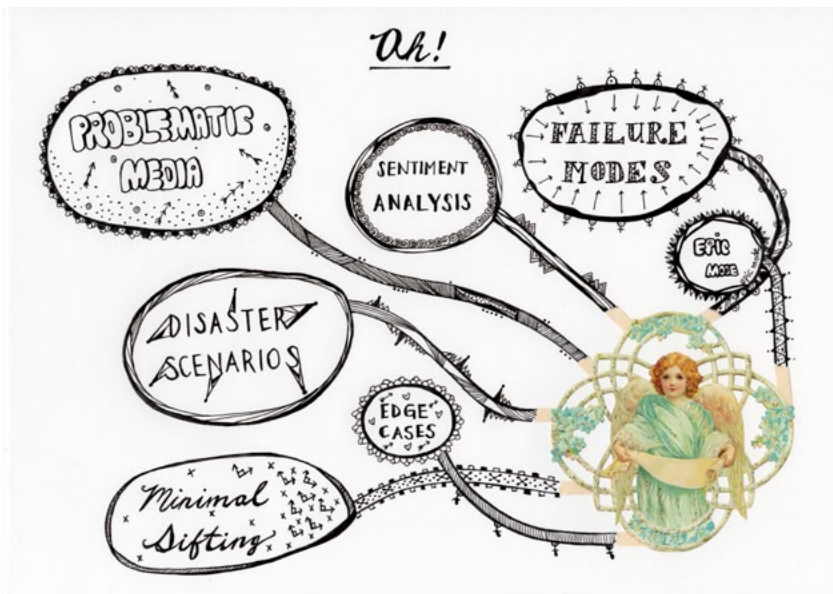


Fig 19.5. YOUR WORK IS OXYGEN FOR THE FUTURE (excerpt) by Jennifer Walshe. ©Jennifer Walshe.

11 TextScores is Emergee [@SuperSuperThank], <https://twitter.com/SuperSuperThank>

Syntax and grammar are disrupted here, and that is the point. I've done multiple performances of these scores, and both the experience of engaging with the score and the results produced are different to other text scores I've worked with.

Markov Chain Generators have their limits, however, and the advent of Deep Learning brings new possibilities. Over the last year I've been working to create a neural network which can write text scores. We're in the middle of the laborious process of taking folders full of PDFs, TIFFs, JPGs and DOCs and transcribing them. This is the grunt work of machine learning – cleaning up and formatting thousands of words of text scores to create a master .txt file of everything we can get our hands on.

Even at this level, a lot of decisions have to be made that will affect the output. What do we do with formatting? With blank lines and paragraph breaks? With fonts and italics and words in bold? What if the text score was designed to be a badge? A t-shirt? A mug? My machine-learning friends tell me that we will eventually get to the point where I'll just throw the folder at the network and it will know what to do. Unfortunately we're not there yet.

Beyond the formatting issues, we have to think about the corpus of text which will be used to train the model. This corpus is not the input – it's used simply to get the network to understand what language is. Many of the researchers I've talked to who make networks to generate text use the Bible as their corpus. I can understand. The Bible is in the public domain, it is easily available as an appropriately formatted file. But think about it lads, just for a second.

I spend hours making a corpus comprised of books by early feminists and Gothic writers. Mary Wollstonecraft. Mary Shelley. Edgar Allen Poe. Anne Radcliffe. It runs for several hours and then crashes the system.

Nonetheless, we are already getting outputs heavy with the directives of Fluxus:

Get a girlfriend! Go to bed everyday and get out in the class. Call yasmin and eat a bit. Speak Chinese! Put your weight on the table at the brig. Take your best friends and get rid of the throne. Tell them you are an idiot. Wake up a banknote and the love for you, and 100%. See concert that i'm very beard. Look after your daddy. Talk about lunch. Keep playin'!

9.

In 2016, Deep Mind, the artificial intelligence division of Google, released WaveNet, a generative model for raw audio.¹² A WaveNet is a convolutional neural network that can model any type of audio. It does this on a sample-by-sample basis. Given that audio recordings typically have at least 16,000 samples per second, the computation involved is significant.

Deep Mind's original blog post shows the structure of a WaveNet. Like any neural network, data enters, moves through hidden layers (this schematic shows three – in reality they could be higher in number), and then an output is produced.

The hidden layers are what interest me here. Machine learning borrows the concept of the neural network from biology. The neural networks in our brains have hidden layers. For example, the networks in our brains relating to sight contain layers of neurons that receive direct input from the world in the form of photons hitting our eyeballs. This input then travels through a series of hidden layers which identify the most salient aspects of what we're seeing, before rendering what we are seeing into a 3D whole.

The Nobel-winning physicist Frank Wilczek describes how 'Hidden layers embody...the idea of emergence. Each hidden layer neuron has a template. It becomes activated, and sends signals of its own to the next layer, precisely when the pattern of information it's receiving from the preceding layer matches (within some tolerance) that template...the neuron defines, and thus creates, a new emergent concept.'

This blows my mind. Hidden layers of calculations. Neurons buried in vast amounts of code, creating and defining emergent concepts. And us, we puny humans, getting to witness new forms of thinking. New forms of art.

12 'WaveNet: A generative model for raw audio' in *Google Deep Mind* (8 September 2016), <https://www.deepmind.com/blog/wavenet-a-generative-model-for-raw-audio>

10.

Generative deep neural networks such as WaveNets can be used to generate any audio – text, voices, music, sound, you name it. It simply depends on what the model is trained on. Think of it like this: a sculptor takes a block of marble, and carves away everything that is not necessary to produce a sculpture of a horse. A generative deep neural network takes a block of white noise, and carves out everything that is not necessary to make the sound of a horse.

In his famous essay ‘The Grain of the Voice’ Roland Barthes writes how ‘the “grain is the body in the voice as it sings, the hand as it writes, the limb as it performs.... I shall not judge a performance according to the rules of interpretation... but according to the image of the body given me.”’ Where is the grain in these recordings? Who does it belong to? How are we to judge something that has no body?

I was stunned when I first heard the WaveNet examples. I’ve learned to reproduce these examples with my voice and have used my versions in pieces such as *IS IT COOL TO TRY HARD NOW?*. I have listened to these examples many, many times. I can hear fragments of a language beyond anything humans speak. I can hear breathing and mouth sounds. I can hear evidence of a biology the machine will never possess. And I know that regardless of its lack of biology, the voice will be read as having gender, ethnicity, class, ability, sexuality. The voice will be judged.

The process of listening to the output of a neural network and trying to embody those results; the experience of trying to take sounds produced by code and feel my way into their logic – for me, that is part of the texture of being a person, of being a composer in 2018.

11.



Generated sound with Walshe’s voice of sounds like hard typing, mumbling, singing, and breathing. Duration: 0:22.

<https://hdl.handle.net/20.500.12434/b9367366>



6th Audio: Jennifer Walshe’s voice, generated by Dadabots, Epoch 27.

CJ Carr and Zack Zukowski, who operate under the name Dadabots, use machine learning to create what they call artificial artists.¹³ They use a modified version of the recurrent neural network SampleRNN to do this (SampleRNN itself was developed in response to WaveNet). CJ and Zack take pre-existing music and use it to train their network, which outputs ‘new’ albums by bands such as the Beatles and Dillinger Escape Plan.

I send CJ and Zack a link to a folder of recordings of my voice. A few weeks later, they send me back a link to a folder containing 341 sound files. 341 sound files of their network learning to sound like me. Hours and hours of material.

Listening to the files is a surreal experience. You should know—you’ve been listening to them over the course of this talk. I find the initial files hilarious—series of long notes, the network warming up in the same way a musician would. As I work my way through the files, they evolve and improve. I can hear myself, and it’s both uncanny and completely natural. The hidden layers have coughed up all sorts of tasty stuff, and I begin to hear certain habits and vocal tics in a new light. I start to hear my voice through the lens of the network.

12.



Generated sound with Walshe’s voice, unrecognizable sounds and breathing. Duration: 0:26.
<https://hdl.handle.net/20.500.12434/6421b917>



7th Audio: Jennifer Walshe’s voice, generated by Dadabots, Epoch 39.

As a performer, my voice operates closely in collaboration with technology. It has been this way ever since I was seven years old and used my First Holy Communion money to buy a tape recorder. It’s rare for me to perform without a microphone, and I’m used to the idea that when I’m performing, my voice exists as multiple overlapping voices, some of which the sound engineer has more control over than me. There are the resonating chambers inside my head; there is my monitor; and there is the PA.

13 Neural networks in *Dadabots*, <https://dadabots.com/>

I'm used to my voice being sampled, vocoded, autotuned, layered with effects. I sing through SM58s and Neumanns, through homemade microphones attached to drum kits. I sing into wax cylinder machines and through exponential horns. I do battle with Max/MSP. I put contact microphones on my throat; I allow the Science Gallery in Dublin to send a transnasal camera down my throat to film my vocal cords. Would I have been a singer, one hundred years ago? Probably, but a very different one. Perhaps, a very frustrated one.

My voice evolved through free improvisation – I have never had any vocal training beyond the discipline I impose on myself as an improviser. The free improvisation duo is one of the dearest relationships to me. The people I've played with the most – Tony Conrad, Panos Ghikas and Tomomi Adachi – are like family members. My strange brothers.

At the moment I'm working on ULTRACHUNK, a collaboration with the artist and machine-learning specialist Memo Akten.¹⁴ I sit in front of my laptop and film myself improvising. Memo uses these recordings to train a convolutional neural network called Grannma (Granular Neural Music & Audio). I think about what it means to improvise for an artificial intelligence system. Who is listening, and how? The goal is that Grannma will be able to produce both audio and video in collaboration with me in a performance situation. No midi, no sampling – a network synthesizing my voice, live. We will witness Grannma learning to see, and learning to listen, live. I'll grapple with performing with a network whose results I cannot predict. A new, strange sibling, both of us together in the uncanny valley on stage.

13.



Walshe's synthesized voice saying "But do I actually want to download all that data. I mean there is a lot of dumb things in a passive aggressive likes and what if they logged all the times my grandma used my account to look at my cousin's profile pics to see if she was not lying about being pregnant." Duration: 0:14.
<https://hdl.handle.net/20.500.12434/022b25a7>



8th Audio: Dubliner quote spoken by version 2 of Walshe's synthesised voice.

14 Memo Akten, Selected Projects in *Memo*, <https://www.memo.tv/works/#selected-works>

In *The Voice in Cinema*, film theorist Michel Chion describes how when we hear a voice in a film soundtrack and ‘cannot yet connect it to a face – we get a special being, a kind of talking and acting shadow to which we attach the name *acousmêtre*.’ Chion discusses what he terms the ‘complete *acousmêtre*, the one who is not-yet-seen, but who remains liable to appear in the visual field at any moment.’ The moment of appearance of the body belonging to the voice results in ‘de-acousmatization’ a process which is ‘like a deflowering.’ Chion attributes awesome powers to the *acousmêtre* – ‘the ability to be everywhere, to see all, to know all, and to have complete power.’ He compares the *acousmêtre* to the voice of God, to the voices of our mothers we heard in the womb before we were born.

The voices that emerge from machine-learning systems, voices that will never be seen, voices without bodies or homes – Chion’s theory of the *acousmêtre* suggests one way to think about them. And as after a prolonged voiceover in a film, I start to wonder when the voice’s body will appear, and what it will look like.

14.



Generated sound with Walshe’s voice of unrecognizable sounds, moments of static like sounds, breathing and some vocalization.

Duration: 0:36.

<https://hdl.handle.net/20.500.12434/6eba0fe6>



9th Audio: Jennifer Walshe’s voice, generated by Dadabots, Epoch 40.

When I was a student, I was taught neat narratives about the development of Western music. The historical inevitability of serialism, the emancipation of the dissonance, the liberation of noise. In terms of the larger picture, we are about to leave all of this in the dust. I am convinced that not only the development of the music, but life in the twenty-first century will be primarily marked by how we engage with, respond to and think about AI. If we care about the world, if we’re curious about human and non-human beings, art and consciousness, we need to be thinking about AI. We have a long way to go, but we can see glimpses of the future already in the world around us – autonomous vehicles, machine-learning-aided cancer diagnosis, neural networks

making accurate predictions of schizophrenia onset, high-frequency trading, gene editing.

What does AI mean for music? Everything. I'm convinced that within fifteen to forty years, machines will be able to write music, in many genres, which is indistinguishable from that written by humans. Well-defined genres with clear rules will be the first to be automated - film music, music for games, advertising music, many genres of pop songs. The machines will write this music more quickly, and more cheaply, than humans ever will. A lot of musicians will be out of work.

Over the next forty years, AI will completely change the way music is made and who it's made by. AI will change the reasons why music is made, and will make us question what the function of music and music-making is. Let me be clear – I don't think humans will ever stop making music. And I think a great deal of music in the future will be made without AI. The challenge of the future will be deciding what it means to make music when the machines can. We will have to think about what it means to make music when, in many cases, the machines will be able to make music of a far higher standard than many humans can.

By engaging with AI, we look at the world in its entirety. We continue asking the questions asked by musicians like the League of Automatic Music Composers, by George Lewis and Voyager.¹⁵ We see how Bach chorales written almost 300 years ago are used by one of the most powerful corporations in the history of the world to train neural networks. We get to be ethnomusicologists romping through the Wild West section of the uncanny valley.

We are all involved, we are all enmeshed, we are all implicated in the development of AI, regardless of whether we code or not, regardless of whether we ever make a piece of music using AI. Every second of every day, our behaviour provides the data for machine-learning systems to train on. We foot the bill for the hardware necessary to do some of these computations – for example, most smartphones now contain dedicated chips for machine learning. Our interactions with our phones – and by that I mean our every waking moment – provides training data for

15 League of Automatic Music Composers (2023), <https://www.dramonline.org/labels/league-of-automatic-music-composers>
 Claude Nobbs, 'George E. Lewis – Voyager Duo 4' clip, YouTube (Aug 3, 2015)
<https://www.youtube.com/watch?v=hO47LiHsFtc>

neural chips on our phones, data that drives the creation of AI at the corporate level. And of course, the AI that develops at the corporate level will be the intellectual property of the corporation. And the AI at the corporate level will define the structure of all of our futures. Every single challenge facing us as a species can either be faced successfully or exacerbated by how we engage with AI.

I am not a computer scientist. I'm a composer who is living in the twenty-first century and trying to think it through. I'm both sublimely excited and blackly horrified about what is coming. I'm trying to give you a sense of how I view the world, and where I think things are going, because that psychological space is where my art comes from. It's a magical space, that is by turns speculative, uncanny and hidden, but most of all deeply embedded in the here and now of the world.

Where do we go from here? AI. What is coming next? AI. What a time to be alive.

Darmstadt, 25 July 2018.