

Speculative Voicing
A Sonic Speculative Design Methodology
for Vocal Imaginaries in the AI Era

Submitted for the Degree of
PhD, School of Communication
at the Royal College of Art

2024

Amina Abbas-Nazari

©Amina Abbas-Nazari [2024, PhD].

This thesis is copyright material and no quotation from it may be published
without proper acknowledgment.

Word Count: 41,707

Speculative Voicing

A Sonic Speculative Design Methodology for Vocal Imaginaries in the AI Era

Voice increasingly mediates artificially intelligent (AI)-enabled communication, with the expanding proliferation of conversational AI systems like Amazon's Echo, voiced by 'Alexa'. This research is concerned with the sound and sounding of voices (human and synthesised) in conversational AI systems and identifies that vocal profiling underpins current understandings of vocal sounding by AI and the AI industry. Vocal profiling relies on normative assumptions that risk misrepresenting individuals, negatively impacting those already marginalised. By fundamentally reorienting the discussion of vocal profiling from 'listening' towards 'sounding', the voice is given increased agency. The practice-led research created vocal imaginaries that reveal and resist vocal profiling, questioning and disrupting the dominant understanding of voice(s) promoted by AI. Situated at the intersection of sound, design and technology, this research incorporates contemporary societal discourse on identity politics, personhood, being and ecology.

Through this PhD research, a sonic speculative design methodology emerged as an original contribution to knowledge, where working with and through sound in the design process created opportunities to establish novel concepts with a turn towards co-creation. Termed 'Speculative Voicing', sonic thinking was successfully applied to speculative design to confront contemporary critique of the field, generating a new materialist, intersectional approach. It was developed over six participatory workshops with young people and then applied to an investigation of vocal profiling in conversational AI systems through two case study projects. Evaluation of the works took place in a further workshop with industry professionals. The evaluation found that the Speculative Voicing methodology accomplished the creation of vocal imaginaries that resist vocal profiling while

creating space to apply narratives that can also reveal them.

Starting from my position as a practising speculative designer and vocal performer, I propose the Speculative Voicing methodology to work with the voice in conversational AI as a design material and treat it as an experimental singer would. Vocal potential is thus generated, building dynamic relations and entanglements to explore concepts of being and identity, supported and evidenced by the motivations of a lineage of female experimental vocalists. Guiding principles of the methodology explorations posit that the voice is always polyphonic through its materiality – co-created with many bodies, environments, matter and the AI system. The recurring metaphor of a vacuum highlights the sonic materiality of voice. Critique is constructed through reflexive analysis of the practice work that demonstrates vocal potentiality in comparison to current profiled presentations and representations of voices in contemporary conversational AI systems.

Speculative Voicing provides a practice-focused, voice-led methodology of working with the polyphonic sonic materiality of voices to reveal and resist voice profiling practices. It is intended for those who want to explore speculative design practice from an intersectional position, especially when working, or intending to work, with voice. The methodology is supplemented by accessible outputs for further impact, dissemination, and advocacy against current practices of vocal profiling. These comprise a Speculative Voicing Framework, a prototyped participatory workshop and two interactive tools to reflect and offer feedback on the understanding, design and implementation of voices in conversational AI.

Declaration

This thesis represents partial submission for the degree of Doctor of Philosophy at the Royal College of Art. I confirm that the work presented here is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. During the period of registered study in which this thesis was prepared the author has not been registered for any other academic award or qualification. The material included in this thesis has not been submitted wholly or in part for any academic award or qualification other than that for which it is now submitted.

Amina Abbas-Nazari, 31st January 2024



Copyright

This text represents the submission for the degree of Doctor of Philosophy at the Royal College of Art. This copy has been supplied for the purpose of research for private study, on the understanding that it is copyright material, and that no quotation from the thesis may be published without proper acknowledgement.

Navigating this Thesis

This is a PhD by Project submission.

This thesis is accompanied by documentation and supporting audio and video materials for the practice projects, in addition to the in-text figures.

In-text hyperlinks denoted by ‘([Item x](#))’ can be found within the text and can be clicked to open relevant materials in relation to the written component.

Alternatively, all these materials can be found in a Google Drive folder (<https://tinyurl.com/373xuer7>) where you can find the Items appropriately titled corresponding to the markings in the written thesis.

[Item 1](#) details the credits, contributors and supporters of all the practice projects. A full index of all the Items follows.

The thesis footnotes are also important. They provide signposting, discursive notes and supplementary references.

Contents:

<u>Abstract</u>	<u>2</u>
<u>Declaration</u>	<u>4</u>
<u>Navigating this Thesis</u>	<u>5</u>
<u>List of ITEMS</u>	<u>10</u>
<u>List of FIGURES</u>	<u>12</u>
<u>Acknowledgements</u>	<u>15</u>
<u>Definition of Terminology</u>	<u>17</u>
<u>Introduction</u>	<u>21</u>
Origins	21
Context	22
Methods	24
Methodology	27
Practice Projects	29
Research Questions	33
Aims	34
Original Contribution to Knowledge	34
Thesis Structure	34
<u>Chapter 1: Literature Review</u>	<u>36</u>
Introduction	36
Literature Gap and Original Contribution	36
Voice Profiling Practices in Conversational AI Systems	38
Profiling Humans from Their Voices	40
Profiling Practices in the Design of Synthesised Voices	42
Voice, AI & Ethics	44
Refusing Representation	45
Voicing in a Vacuum	48
Voicing Materiality in Practice	53
Voicing Many Voices	54
Conclusion	56
<u>Chapter 2: Voice as Sonic Material</u>	<u>58</u>
Vocal Practice to Practice-Led Research	58
Choirs	58
Solo Vocal practice	61

Trying to teach an AI to Sing	63
The Voice...Sometimes Behaves So Strangely	65
Conclusion	67

Chapter 3: Participatory Workshops - Developing a Sonic Speculative

<u>Design Methodology</u>	69
Introduction	69
Speculative Listening Workshop Structure and Process	70
Speculative Listening I	72
Speculative Listening II	76
Speculative Listening III	80
Multiphonic Connections	83
Giving Voice to Synthetic Sonics	87
Speculative Listening IV Workshop	91
Final Reflections and Conclusion	93

Chapter 4: Methodology – Towards Speculative Voicing

Introduction	95
Introduction to Speculative Voicing	96
Speculative Design meets Sonic Thinking	98
Developing Speculative Voicing as a Methodology	101
Polyphonic Materiality in Practice	104
A Speculative Voicing Framework	109
Conclusion	115

Chapter 5: Polyphonic Embodiment(s)

Introduction	116
Project Origins	116
Human Voices and Practices of Profiling	118
Notes on the Dataset	121
Recreating the AI	123
Human Voices in a Vacuum	125
Human Voice as Material	126
Making and Using the Devices with our AI	129
Speculatively Voicing Human Voices in Conversational AI Systems	135
Analysis of Polyphonic Embodiment(s)	140
Conclusion	143

Chapter 6: Acoustic Ecology of an AI System

Introduction	146
--------------	-----

Notes on the Project Title / Iterations and Development	146
Text-to-Speech, Synthesised Voices and Practices of Profiling	147
Synthesised Voices in a Vacuum	149
Synthesised Voice as Material	152
Making of Acoustic Ecology of an AI System and Technical	
Considerations	153
Speculatively Voicing Synthesised Voices in Conversational	
AI Systems	159
Analysis of Acoustic Ecology of an AI System	163
Conclusion	165
<u>Chapter 7: Evaluation of Speculative Voicing</u>	167
Introduction	167
Workshop Structure and Process	167
Workshop Participation and Participants	169
Evaluation of Polyphonic Embodiment(s)	170
Evaluation of Acoustic Ecology of an AI System	173
Analysis of Findings	175
Workshop Discussion and Wider Implications	177
Conclusion	179
<u>Chapter 8: Conclusion</u>	182
Introduction	182
Outputs	182
Speculative Voicing Workshop (for AI Industry Professionals)	182
Speculative Voicing Framework (for Speculative Designers and	
Advocacy Groups)	189
Interactive Tools (for Speculative Designers and Non-Specialists)	189
Documentation (for the Public)	192
Findings	193
Limitations and Considerations	195
Original Contributions to Knowledge	196
Future Work	197
Advocacy Work	197
Theoretical Work	197
Practice	198
New Challenges	198
<u>Appendices</u>	200
Appendix A Inaudible Audio Track Listing	200
Appendix B: Multiphonic Connections Automated Telephone Script	202

Appendix C: Mural Board from IBM Workshop	206
Appendix D: IBM Workshop Google Form Responses	207
Appendix E: Public Engagements Since Commencing PhD Study	209
<u>References</u>	<u>211</u>

List of ITEMS

'Items' are audio, audio visual, or accompanying media that contribute to this thesis. They are available at the accompanying in-text links and / or via: <https://tinyurl.com/373xuer7>. Unless otherwise stated, all media was created by the author.

Items:

1. Practice projects credits, contributors and supporters
2. Trying to teach an AI to Sing AI synthesised clone of my voice, example 1
3. Trying to teach an AI to Sing AI synthesised clone of my voice, example 2
4. Trying to teach an AI to Sing experiment, example 1
5. Trying to teach an AI to Sing experiment, example 2
6. The Voice...Sometimes Behaves So Strangely audio track
7. Speculative Listening workshops 'Inaudible Audio' soundtrack tracklist.
8. Multiphonic Connections voicemail message
9. Multiphonic Connections voicemail message
10. Multiphonic Connections voicemail message
11. Multiphonic Connections voicemail message
12. Multiphonic Connections voicemail message
13. Multiphonic Connections voicemail message
14. Multiphonic Connections voicemail message
15. Multiphonic Connections automated telephone introductory message
16. Multiphonic Connections automated telephone initial listening exercise
17. Multiphonic Connections automated telephone voicemail instructions
18. Giving Voice to Synthetic Sonics preview of improvised performance
19. Google Colab wav2face documentation of application in use. Sitraka Rakotoniaina, Amina Abbas-Nazari & Nestor Pestana
20. Polyphonic Embodiment(s) video documentation of material experiments. Amina Abbas-Nazari & Nestor Pestana
21. Polyphonic Embodiment(s) audio clips inputted into Google Colab wav2face
22. Polyphonic Embodiment(s) video documentation of devices and AI generated image. Amina Abbas-Nazari & Nestor Pestana
23. Polyphonic Embodiment(s) audio clip of Device #1
24. Acoustic Ecology of an AI System seven audio clips from work
25. Acoustic Ecology of an AI System video documentation preview on the Research and Waves platform
26. speaker-2 Google 'WaveNet Babble' audio clip. (Oord & Dieleman, 2016)
27. 2 'smelting' audio clip from Acoustic Ecology of an AI System
28. speaker-5 Google 'WaveNet Babble' audio clip. (Oord & Dieleman, 2016)
29. IBM Workshop Participant descriptions of audio, noted on Mural
30. Documentation of Speculative Voicing Workshop at the Articulating Data Symposium. Euan Gilmour, Video Production Edinburgh.
31. Voicing Beyond the Vacuum Max/MSP patch prototype
32. Voicing Beyond the Vacuum Max/MSP patch preview

33. Voicing Beyond the Vacuum Max/MSP Patch (for download and use). Andy Sheen & Amina Abbas-Nazari

List of FIGURES

Unless otherwise stated, all figures, images and diagrams were created by the author.

Figures:

1. Conversational AI Systems. Adapted from (Anon, n.d.).
2. Voice and speech recognition in conversational AI systems. Adapted from (Anon, n.d.).
3. Reciprocated human and synthesised voice profiling attributes.
4. Practice process / thesis structure diagram.
5. PhD Summary table.
6. International Prototype Kilogram (IPK). National Institute of Standards and Technology, US / Public Domain.
7. Materials for Speculative Listening workshops. Lisa Marie Bengtsson.
8. Life Rewired Hub, Barbican Centre, sectioned off with semi-transparent curtain.
9. Speculative Listening device made during Barbican workshop. Lisa Marie Bengtsson.
10. Using the silent disco headphones and microphone during Barbican Speculative Listening workshop. Lisa Marie Bengtsson.
11. View of workshop taking place at Barbican, from above. Lisa Marie Bengtsson.
12. Children taking it in turns to present their devices. Lisa Marie Bengtsson.
13. A Speculative Listening device to hear the sound of conception - when a sperm meets the egg for sex education purposes. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys, all rights reserved.
14. A Speculative Listening device listening to deep underground sounds via the hip bones. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys, all rights reserved.
15. Possible conditions that limit humans' capacity to hear certain sounds, Speculative Listening II workshop. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys, all rights reserved.
16. A Speculative Listening device to listen to the menstrual cycle to better understand changes in mood and communicate this to others.
17. A Speculative Listening device for the sonification of chaos and personal decision making activated by the breath.
18. A Speculative Listening device to sonically describe anxiety (emanating from the stomach).
19. Multiphonic Connections social media launch poster.
20. Multiphonic Connections social media poster with telephone number.

21. Abbreviated voicemail message from Multiphonic Connections.
22. Abbreviated voicemail message from Multiphonic Connections.
23. Poster for Giving Voice to Synthetic Sonics. Features schematic from (Dudley, 1940)
24. Participant blowing into plastic tube from Speculative Listening I. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys, all rights reserved.
25. Participants listening to 'Inaudible Audio' soundtrack. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys, all rights reserved.
26. Participant making their Speculative Listening device to transform tactile interaction into sound from Speculative Listening IV. Tim C Huang.
27. An Unresolved Mapping of Speculative Design V2.0. (Montgomery, n.d.)
28. Speculative Voicing in An Unresolved Mapping of Speculative Design V2.0. Adapted from (Montgomery, n.d.)
29. Vibration sensor.
30. Vibration sensor, wired up.
31. Speculative Voicing Framework schematic.
32. Initial sketch ideas for Polyphonic Embodiment(s). Amina Abbas-Nazari & Nestor Pestana.
33. Initial sketch ideas for Polyphonic Embodiment(s). Amina Abbas-Nazari & Nestor Pestana.
34. Testing wav2face Google Colab.
35. Timed spectrogram test for Polyphonic Embodiment(s).
36. Device #4 from Polyphonic Embodiment(s). Amina Abbas-Nazari & Nestor Pestana.
37. Polyphonic Embodiment(s) table of vocal qualities, effect and materials used for device.
38. Device #1 from Polyphonic Embodiment(s). Amina Abbas-Nazari & Nestor Pestana.
39. Example wav2face AI generated face.
40. Ray Space audio plug-in software screenshot. (QuikQuak, n.d.).
41. Crowd Chamber audio plug-in software screenshot. (QuikQuak, n.d.).
42. Acoustic Ecology of an AI System table of Audio Title, Locations on Map, Voice Sound and Sounding Qualities table.
43. Screenshot of home page of Acoustic Ecology of an AI System on Attune / Research and Waves website. (Abbas-Nazari, 2020).
44. IBM Workshop Schedule.
45. Example evidence of Speculative Voicing methodology from IBM Workshop evaluation.
46. Speculative Voicing Workshop. Elspeth Murray.
47. Voice Quality descriptor terms. Adapted from (Singh, 2019, p. 242).
48. Speculative Voicing Workshop: strangled. Elspeth Murray.
49. Speculative Voicing Workshop: strangled. Elspeth Murray.

50. Speculative Voicing Workshop: modal. Elspeth Murray.
51. Speculative Voicing Workshop: broad. Elspeth Murray.
52. Speculative Voicing Workshop: low. Elspeth Murray.
53. Speculative Voicing Workshop: intense. Elspeth Murray
54. Screenshot of video documenting prototype version of Max/MSP patch
55. Screenshot of final Max/MSP patch. Amina Abbas-Nazari & Andy Sheen
56. Screenshot of Speculative Voicing Website.

Acknowledgements

This research was supported by Techne (NPIF), who provided funding, without which I could not have undertaken this work. I am also grateful for the brilliant training they arranged, the opportunities they created to meet other doctoral students, and the team's reassuring presence during the uncertainty of early COVID-19 lockdowns.

There have certainly been difficult moments during this PhD, but I am incredibly grateful for the privilege of being given this opportunity, which I have greatly enjoyed. Many people have helped to bring this PhD to fruition and guided me on my research journey.

In addition to the collaborators and supporters noted in the thesis, I would like to thank:

My supervisors, Luke Pendrell and Dr Matt Lewis, for taking me and my project on, with your open-minded guidance and support.

My PhD examiners Dr Annie Goh & Prof. Andrew Knight-Hill for their rigorous questioning, engagement and encouragement of my work.

School of Communication Postgraduate Lead, Prof. Teal Triggs, who has always been an email away. Thank you, Teal, for helping me through the undulations of PhD study and facilitating our wonderful weekly Methods of Intent classes.

Our programme administrator, Caroline Vulela, for keeping everything organised and ensuring we always have the necessary information.

My School of Communication postgraduate peers for their rich discussion and convivial conversation.

I've also had a lot of help from people to advance the skills needed to complete a PhD. For this, I thank Sarah Fitzalan Howard and English for Academic purposes tutor, Dr Sarah Blair, for their patience and kindness, and Cathy Johns for proofreading this thesis.

Prof. Johnny Golding, for allowing me to join the School of Arts and Humanities Entanglement reading group in my earlier years of study.

Dr Helga Schmid and Dr David Chatting for their feedback during the PhD application process and Dr Wesley Goatley for being a critical friend during viva preparation.

Prof. Tony Dunne, Fiona Raby, and Nina Pope, for the freedom and encouragement during my MA studies to start exploring some of the themes in this thesis.

Finally, so many thank-yous to my partner, Aimee. Thank you for so graciously supporting me through all the challenges that PhD study has brought and for helping me find my voice. And, of course, thanks to our little cat Sodamaus for (quite literally) being by my side during my studies.

Definition of Terminology

Conversational AI Systems – Describes the technology behind automated speech-enabled applications that offer human-like interactions between computers and humans (Anon, n.d.). A typical case of this is human voice interaction with an Amazon Echo device, voiced by the synthesised voice ‘Alexa’, and this is used as a recurring example in this thesis. These voice-mediated interactions are also known by terms such as ‘voice user interfaces’ (VUIs) (Nass & Brave, 2005). ‘Conversational AI system’ was chosen as the terminology for this thesis because it allows for the discussion of human and synthesised voices and the ‘conversation’ created between them in analysing how they are used, understood and utilised by the AI industry.

Imaginaries – ‘Sociotechnical imaginaries’ are ‘collectively held and institutionally stabilized’ (Jasanoff, 2015, p. 4); ‘imagined forms of social life and social order reflected in the design and fulfilment of nation-specific scientific and/or technological projects’ (Jasanoff & Kim, 2009, p. 120). Kang (2022) examines this dynamic in the AI voice identification and analysis industry, critiquing it as a ‘biometric imaginary’ whereby the voice is reduced to a ‘sound object’ to affix it to identity. However, Mager and Katzenbach (2021) point out that imaginaries can be ‘multiple’ and ‘contested’. In this PhD project, I use my practice to illustrate ‘alternative social imaginaries that open new perspectives’ (Dunne & Raby, 2014, p. 189) that reveal and resist vocal profiling by AI.

Intersectional – Black feminist legal scholar Kimberlé Crenshaw first proposed the term ‘intersectional’ in her 1989 article ‘Demarginalizing the Intersection of Race and Sex’. The term describes the interconnected nature of social categorisations such as race, class, and gender as they apply to a given individual or group. It is regarded as creating overlapping and interdependent systems of discrimination or disadvantage (Crenshaw, 1989). In the context of this thesis, the term is used

independently but also embedded within new materialist theory to incorporate the 'interconnected', entangled nature of humans with non-human entities, matter and other materials.

Machine Listening – An area of research within AI which uses machine learning and signal processing to teach computers to understand and interpret audio data and extract useful information from sound (Parker & Dockray, 2023).

Materiality - The use of the word materiality in this thesis is two-fold. It references new materiality as a philosophical theory and where the voice as a material converges with this ontological stance. Therefore the materiality of voices comprehends the voices as a material, which can then be shaped and formed through convergence and co-creation with other materials, possessing their own material properties.

Normative – Relating to, or deriving from, a standard, typical or average, especially regarding behaviour (Oxford Languages, n.d.)

Ocularcentric – Ocularcentrism is 'A perceptual and epistemological bias ranking vision over other senses in Western cultures' (Oxford Reference, n.d.).

Polyphonic – In music, polyphony means the simultaneous combination of two or more tones or melodic lines, derived from the Greek word for 'many sounds' (*Encyclopedia Britannica*, n.d.). In the context of this thesis, it is used to mean many voices.

Practice-Led Research – The 'primary focus of the research is to advance knowledge about practice, or to advance knowledge within practice. Such research includes practice as an integral part of its method' (Candy, 2006).

Profiling – Profiling is the act or process of extrapolating information about a person based on known traits, tendencies, observed characteristics or behaviour (Merriam-Webster Dictionary, n.d.). In this research, profiling is described in relation to its use in AI to extrapolate vocal sonic data of human traits and tendencies to determine people's characteristics and behaviour. This thesis explores both the use of AI in profiling humans based upon the sound of their voices and profiling imagined human personas for the sound design of AI synthesised voices.

Sounding – As with 'voicing', 'sounding' should be considered a verb and not a noun (see below). This is also iterated by Julian Henriques' (2011) theorising of 'sounding': in his enquiry he connected it specifically with Jamaican reggae sound systems and dancehall culture. As Henriques says, and as this thesis understands, sounding is embodied, materialist and situated. Thinking and working through and with sounding 'serves to draw attention to a rather different object of enquiry than the conventional ones of text or image' and is 'not entirely bound up with language, notation and representation' (Henriques, 2011, p. 2).

Speculative Design - A design practice which provides informed, hypothetical extrapolations of an emerging technology's development with a deep consideration of the cultural landscape into which it might be deployed. Projects speculate on alternative products, systems and worlds by applying different ideologies or configurations to those currently directing product development (Auger, 2013). In this thesis I developed a sonic-centric speculative design to apply an intersectional new materialist ideology to conversational AI technology. I use 'speculative designers' to describe a manner of conducting creative practice which is akin to speculative design, not necessarily a particular type of practitioner or profession.

Vococentric – A term used in cinema and film studies that means to privilege the

voice over other audio (and visual) media (Chion, 1994).

Voice - Schlichter and Eidsheim (2014) note how discussion about the voice occupies many disciplines yet there is little shared terminology. In this thesis I utilise theory from sound and music performance practice to explore the way AI and the AI industry currently comprehends voice. The friction between these modes of understanding is used to generate critique of voice profiling practices in AI. Predominantly, my use of the word voice refers to the sound and sounding produced when voicing, not the linguistic content of speech.

Voicing – The term ‘voicing’ should be considered a verb, not a noun (Eidsheim, 2015, pp. 2-3). It is the voice in action, inter-action, part of an event and state of being. It adopts a new materialist approach to its understanding (Eidsheim, 2015).

Synthesised Voice - An artificially produced human voice created by a computer system (Collins, 2023).

Vacuum - A volume with the absence of matter including air. In this thesis I utilise the metaphor of a vacuum to describe the conditions voices are computed and comprehended as within conversational AI systems. The metaphor draws attention to AI’s lack of appreciation for the voice as sonic and as having sonic materiality in interaction and co-creation with other matter. As Eidsheim (2015) says, ‘sound does not exist in a vacuum but is materially dependent’ (p. 49).

Introduction

Origins

This PhD is driven by a longstanding personal interest in exploring the voice in relation to emerging technology and was largely initiated through my background as a singer, as well as my study of speculative design. I have researched the voice in conjunction with emerging technology through practice since 2008 (See: Abbas-Nazari, 2022). Additionally, this PhD research is undoubtedly influenced by my experiences as a mixed-race, non-binary, queer person. All these identifying terms (which I reluctantly and rarely use) are noted here because they signify experiences of falling between or beyond normative categorising boxes of identity. I find the language of categorisation inadequate, limiting and restrictive in the expression and exploration of identity and being. This view is likely to have been a significant driving factor for the contemplation of vocal profiling in this thesis. Rather than trying either to find a label to contain myself within, generate a new one or straddle multiple boxes, I believe a wholeness can be found in multiplicity, especially when co-created, but not compared to others. As such, this research takes an intersectional position (Crenshaw, 1989). For me, the analogy of polyphony and my experiences of singing in a choir can represent this, which can be seen in this PhD research and practice.

This research is perhaps also subtly informed by my grandmother's role as a telephone exchange switchboard operator, in the years just before telephone systems were automated, in 1960 (See: Science Museum, n.d.). She was the 'voice' of an early form of telecommunication, which set in motion the artificially enabled conversational systems available today. She also received elocution training to develop a 'telephone voice' which was deemed a prerequisite and mandatory to use in the job.

Context¹

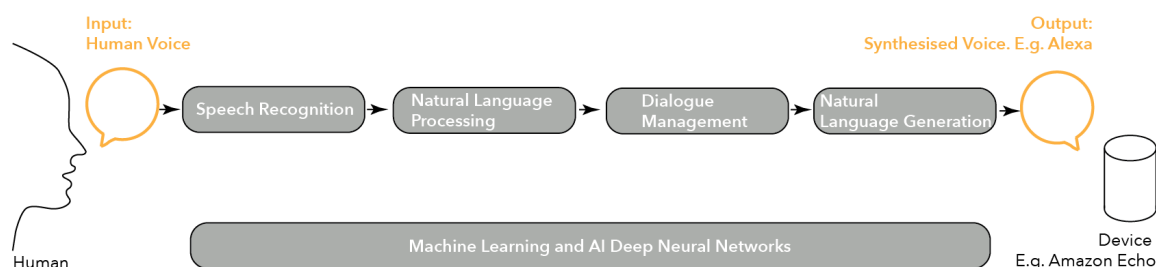


Figure 1: Conversational AI Systems. Adapted from (Anon, n.d.).

Artificial Intelligence (AI) increasingly mediates communication, with the expanding proliferation of networked artificial intelligence-enabled devices in this ‘large-scale’ era of AI.² Conversational AI systems offer voice interactions between humans and computers enabled by AI (Figure 1). Speech recognition systems date back to 1952, with Bell Laboratories’ ‘Audrey’ system (Cox, 2019, p. 215), and the creation of electronically synthesised voices started in 1939 with ‘Pedro’ the Voder (p. 171).³ These technologies are becoming more prominent in the Western world with the increasing use of conversational AI systems.⁴ The urgency of this PhD enquiry is evidenced by Amazon’s announcement in May 2023 that over half a billion Alexa-enabled devices had been purchased globally (Amazon, 2023).

Voice is a multifarious term with differing understandings depending on the field of research. I utilise this complexity in the development of my research by challenging the way AI understands voice with ideas from sound and music practice. I identified that there is a severing between speech and voice that occurs in

¹ The context of this research (voice profiling in conversational AI systems) is a constantly evolving field with vast interest and investment currently. Indeed, much has changed since I started my PhD in September 2018. At times, it has been challenging to stay abreast of developments and research in the field and neighbouring areas of study. I believe all information is correct and accurate at the time of submission and hope that my research will find continued relevance in the years to come.

² Sevilla et al. (2022) note that since 2015, there has been a new, much more dominant AI era with large-scale models and datasets, such as Chat GTP, which have been doubling in growth roughly every 9.9 months.

³ Earlier mechanical devices that synthesised speech date back to the 18th century, with Wolfgang von Kempelen's speaking machine (Cox, 2019, p. 168).

⁴ West, Kraut and Chew (2019) report US market research revealed that 15 million people owned three or more devices equipped with conversational AI in 2018, up from 8 million in 2017 (pp. 92-93).

conversational AI systems. ‘Speech’ recognition is concerned with the linguistic content of spoken information (Singh, 2019, p. 3), and ‘voice’ recognition intends to identify a user by the sound of their voice (Tate, n.d.). My research is most concerned with the input and output of conversational AI systems (Figure 2) – specifically, the sound and sounding of human and synthesised voices, respectively, and with exploring their contemporary understanding. In researching this, I found that they are understood, defined and described through profiling, and this is the case for both human and synthesised voices.

Profiling of voices by AI involves analysing vocal sonic data to determine people’s characteristics and behaviour. Singh (2019) describes how voice profiling aided by AI can present a ‘complete picture’⁵ of an individual from their voice (p. 325). This includes physical (p. 86), physiological (p. 96), demographic (p. 99), medical (p. 101), psychological (p. 104), behavioural (p. 115), and sociological features (p. 116). Similarly, in the case of synthesised voices, imaginary humans, or personas, are profiled to determine their designed sound, aiming to imitate human voices and communication (Nass & Brave, 2005). However, this profiling is limited by its portrayal of very narrow representations of what it is to be human (West, Kraut & Chew, 2019).

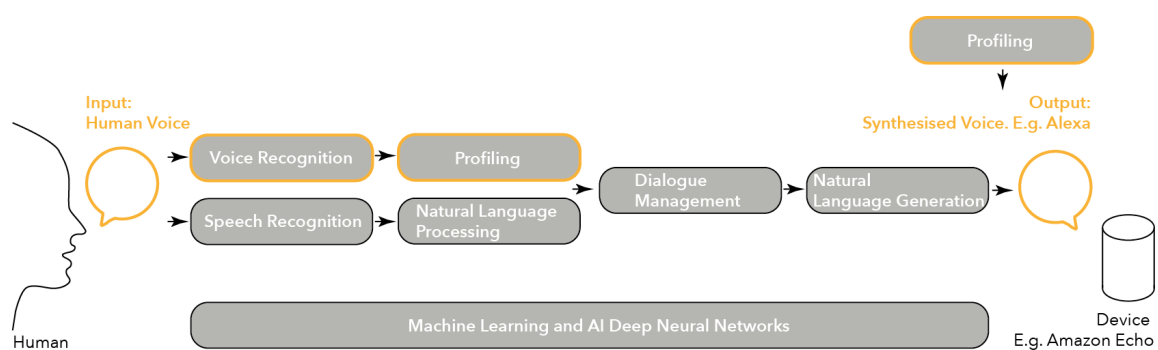


Figure 2: Voice and speech recognition in conversational AI systems. Adapted from (Anon, n.d.).

⁵ I use the phrase ‘complete picture’ to signify the fullness and wholeness that voice profiling intends to achieve. Additionally, Singh (2019) describes how to profile personality traits but also visual information about an individual in Chapter 9 of their book, titled ‘Reconstruction of the Human Persona in 3D from Voice, and its Reverse’ (p. 325).

While AI-enabled facial recognition and profiling have received notable criticism (Buolamwini & Gebru, 2018), resulting in moratoriums on their use in policing and law enforcement by major companies, including IBM, Apple and Amazon in the US (Hao, 2020 b), the same cannot be said for voice recognition and profiling. This research aims to expand the critique of voice profiling, by and within AI, as normative and marginalising (Amaro, 2019; Birhane, 2021) by utilising more holistic understandings of voice from the fields of sound and music practice. This practice-led perspective of vocal sounding suggests that vocal profiling results in oversimplistic assumptions, which are ultimately untenable. Vocal profiling constrains and constricts polyphonic vocal potential, neglecting the material world that the sound and sounding of voices operate within. In my research, I argue that female experimental vocal practitioners evidence this phenomenon, and I frame these vocalists as setting precedents for my research.⁶ The prevailing essence of this research is that of emancipation – identifying openings for new possibilities within existing structures, disrupting systems and challenging pre-defined expectations.

Methods

Practice Methods

As will become apparent, there is a significant cross-over and duplication of my core method and the over-arching methodology of this research, which adopts a sonic-centric speculative design approach, whereby sound and sonic thinking (Voegelin, 2014) meet speculative design (Dunne & Raby, 2014). Speculative design is a design practice which provides informed, hypothetical extrapolations of an emerging technology's development with a deep consideration of the cultural landscape into which it might be deployed. Projects speculate on alternative products, systems and worlds by applying different ideologies or configurations to those currently directing product development (Auger, 2013).

⁶ The vocalists included in this research are Cathy Berberian, Laurie Anderson, Holly Herndon, Pauline Oliveros, Meara O'Reilly, Jennifer Walshe, Elaine Mitchener.

Sound is used as a primary medium or conduit in the speculative design works. I also employ participatory design methods during workshops devised to develop my Speculative Voicing methodology in collaboration with young people.⁷ Here, co-creation as ‘collective creativity’ in the development process (Sanders & Stappers, 2008) shaped the PhD project through more democratic means (Nygaard, 1990; Preece et al., 2015). I also value participatory design as a method to invite people to explore ideas and concepts that they may not have yet had the opportunity to discuss. This case covers conversational AI themes, voices, machine listening and sounding. My work seeks to be accessible and open to those engaged in its process or who will encounter it. Therefore, I use widely available tools, materials,⁸ and open-source technologies in my practice. Whilst this research sits in the context of AI, I do not focus on employing methods or practices that use AI. This bottom-up, do-it-yourself (DIY), democratic approach contrasts with the top-down imposition of most AI systems. It seeks new ways to reveal and resist vocal profiling to question its fundamental rationale.

The thesis research and practice have developed through ‘reflexive practice’ (Schön, 2016), whereby I have enacted ‘reflection-on-action’ of my vocal and design practice to ground prior experiences through theory. Then, in creating new practice works, I employed ‘reflection-in-action’ (Schön, 2016) to provide written self-reflective analysis of my practice being used to test and evolve existing theory.

Analysis Methods

Analysis of the works and workshops use ‘thick description’ – summarised as aiming to ‘capture the thoughts, emotions, and web of social interaction among observed participants in their operating context’ (Ponterotto, 2006, p. 242) to create

⁷ See Chapter 3 for more details.

⁸ It is also intended to have a low environmental impact, using things from around the home, found in the recycling bin, easily recycled or used again.

an autoethnographic account of my practice and the processes used. These writing sections are described as autoethnographic, as I present a personal narrative, description and analysis of the intentions of my practice as its creator. As described by Adams, Ellis, and Jones (2017), 'Autoethnographers believe that personal experience is infused with political/cultural norms and expectations, and they engage in rigorous self-reflection – typically referred to as 'reflexivity' – in order to identify and interrogate the intersections between the self and social life'. Within these sections of text, I aim to write critically and evocatively, to reflect on my work and ignite the reader's imagination. Van Leeuwen (2016) discusses the multi-modal nature of media, highlighting how descriptive qualities traverse different mediums and could be considered a social semiotic theory of synaesthesia. For example, listening to and describing sound evokes a multi-sensorial consideration of qualities such as colour, texture, and temperature, even though sound does not possess these attributes. In the text-based analysis of my sonic media-based case studies, I use written language to simulate multi-sensorial states of comprehension, aligning with the intentions of the sound works produced.

It is important to note the subjectivity embedded within my work, especially in how I design the sound and sounding of voices within my projects from my own perspective, understanding and previous life experiences. In addition, the way people interpret sound, and therefore my work, is 'highly dependent on context and var[ies] across time and space', despite 'remarkable continuity in the cultural connotations of particular sounds' (Franinovic & Serafin, 2013, p. 5). Therefore, the understanding of my work is co-enacted between me and its viewers or audience. Here, I want to draw attention to the way that my case study projects are not intended to define how human and synthesised voices *should* sound in conversational AI systems. Instead, I am proposing a methodology of working and understanding, by which the projects present open-ended examples of how voices *could* be implemented. The methodology could be applied by someone else, for

example, and the outcome would be different, especially as sound is an open-ended, evocative medium to access people's imagination (Voegelin, 2014). This, in turn, also foregrounds the polyphonic potential of voice.

Methodology

In my practice-led research (Candy, 2006), I developed and defined Speculative Voicing as a methodology to produce vocal imaginaries that resist and reveal voice profiling in conversational AI systems. The sonic speculative design methodology is ultimately vococentric, in the case of this research, prioritising the voice over any other audio or visual material (Chion, 1994). This vococentric position shifts the contemporary discussion of vocal profiling from (machine) listening towards (vocal) sounding, giving the voice precedence and increased agency.⁹ I foreground vocal sounding to contest the way that voices are currently heard and voiced by AI systems. This methodology applies sonic thinking to speculative design practice, whereby sounding is a form of speculative imaginary (Voegelin, 2014). Furthermore, I explore how the voice can become, or is becoming, a design material, especially now that conversational AI systems are increasingly prominent in everyday lives. Findings generated from the exploration of this methodology build upon the field of speculative design to address its contemporary critique via an intersectional position.

⁹ See Chapters 1 and 2 for further discussion.

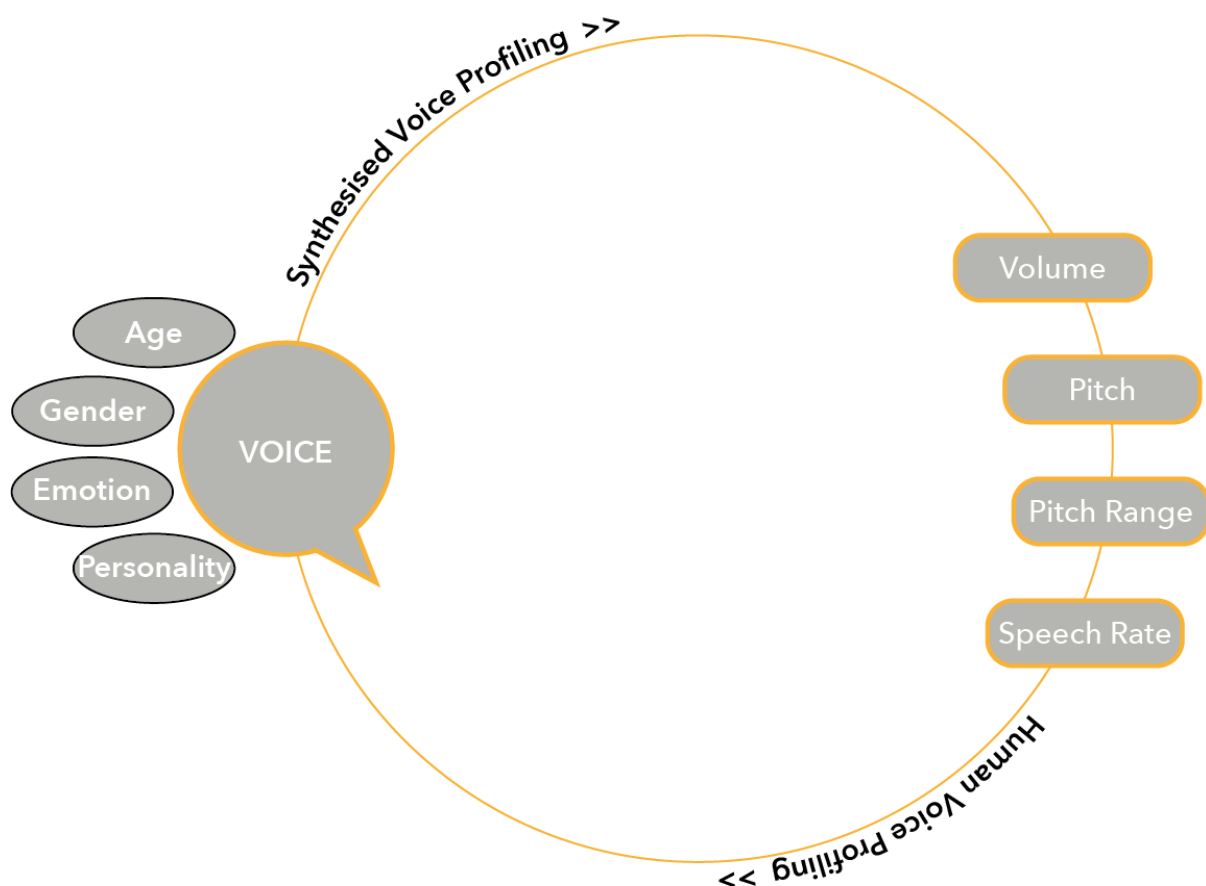


Figure 3: Reciprocated human and synthesised voice profiling attributes.

Theory articulated in Eidsheim's (2015) book *Sensing Sound: Singing and Listening as Vibrational Practice* provides the basis of my methodology by describing the material dependency that voice has in its construction and how it is co-created with other matter. For example, the sounding of voices is different in air and in water. The materiality of voices comprehends the voices as a material, which can then be shaped and formed through convergence and co-creation with other materials, possessing their own material qualities. In the thesis I use a reoccurring metaphor of a 'vacuum' to further highlight vocal materiality. I develop and orientate Eidsheim's theories to enable the investigation of the materiality of voices in conversational AI systems. I define a Speculative Voicing Framework, consisting of four conditions, seeking to explore and understand the polyphonic potential of voices in these systems.

Voices are currently profiled through four auditory attributes: volume, pitch, pitch range and speech rate. By amalgamating research from different sources (Nass & Brave, 2005, pp. 34-36; Feldman, 2016; Singh, 2019, pp. 4-5), I deduced that this current framework acts in a reciprocal and reinforcing way between understandings of human voices and the design of synthesised voices by and within AI (Figure 3). However, Speculative Voicing looks beyond voice profiling to explore the sound and sounding of voices shaped by the body, the environment, other matter, bodies, and the conversational AI system itself, seeking to break this reinforcing and reciprocal cycle of understanding employed in vocal profiling. In turn, Speculative Voicing intends to explore a more holistic understanding of voice from an intersectional, social and ecological standpoint. The four conditions I stipulate provide an alternative framework for understanding and presenting voices in conversational AI, which function to reveal and resist vocal profiling. With this, I explore how one voice in conversational AI systems can:

1. be embodied and co-created with other matter and/or the environment
2. be embodied and co-created with many other bodies and voices, such as in a choir
3. embody many voices, co-created with the body
4. embody many disembodied voices, co-created with the conversational AI system

Practice Projects

By its nature, a project concerning the materiality of voices requires a practice-led approach – to enact and activate their sound and sounding. Every stage and chapter of the thesis is accompanied by practice projects which direct the research towards answering the thesis questions. [Item 1](#) contains full acknowledgement of contributors and supporters for all the practice works. The projects operate as stand-

alone works but also guide the research narrative, as shown in Figure 4.¹⁰ The dashed or solid line stroke style surrounding each practice component denotes which of the thesis questions they helped to answer. The case study practice projects are intended as open-ended illustrations of vocal imaginaries and explore opportunities for alternative narratives beyond profiling practices through voice. The speculative nature of all the works/workshops demonstrates a polyphony of ideas which are, as Dunne (2009) says, ‘facilitated by design, not determined by it’.

For all the workshops that involved participants, I sought the individuals’ consent to be part of this PhD research via an ethics approval procedure, which was overseen and approved by the Royal College of Art’s Research Ethics Committee.

¹⁰ I was inspired to make this diagram after reading Helga Schmid’s (2017) doctoral thesis, which used a similar mapping exercise to plot her practice in relation to the thesis questions.

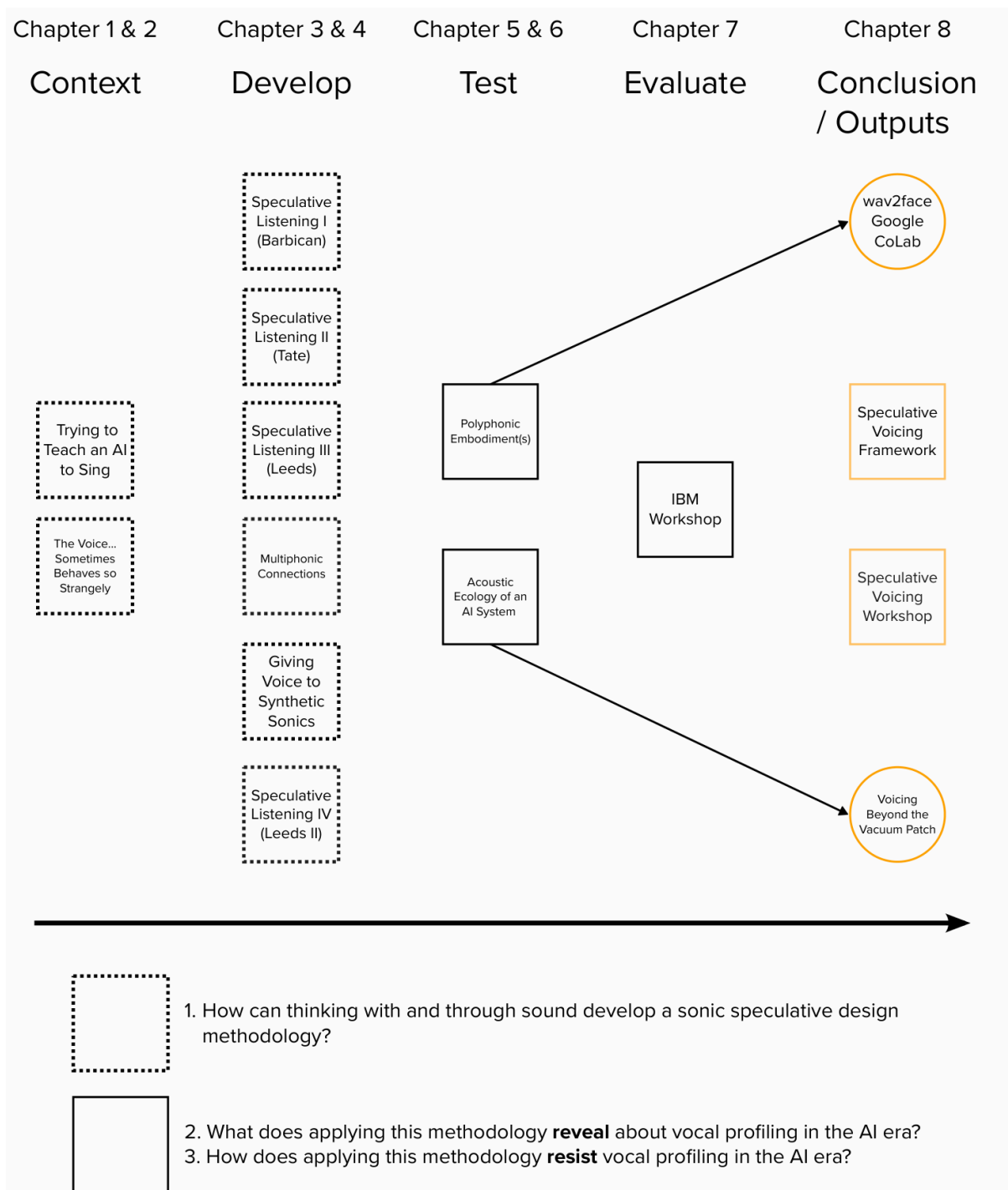


Figure 4: Practice process / thesis structure diagram.

In this practice-led PhD investigation, there are two experiments that position and contextualise my research:

Trying to Teach an AI to Sing

The Voice...Sometimes Behaves so Strangely

A series of six participatory works/workshops that develop my sonic speculative design methodology:

Speculative Listening I - Barbican

Speculative Listening II - Tate

Speculative Listening III - Leeds

Multiphonic Connections

Giving Voice to Synthetic Sonics

Speculative Listening IV - Leeds II

Two case study projects apply and test my methodology:¹¹

Polyphonic Embodiment(s)

Acoustic Ecology of an AI System

One IBM Workshop with its employees is used to evaluate these two case study projects and the methodology.

A final provisional workshop concludes the PhD research:

Speculative Voicing Workshop

Two interactive tools and a framework for future work, dissemination and impact accompany the research:

wav2face Google Colab

Voicing Beyond the Vacuum Max/MSP Patch

Speculative Voicing Framework

Additionally, two publicly presented outputs have been created:

Speculative Voicing Webpage (Abbas-Nazari, 2022)

¹¹ The case study practice project Polyphonic Embodiment(s) investigates the (input) sound and sounding of human voices in conversational AI systems. The other, Acoustic Ecology of an AI System, investigates the (output) sound and sounding of synthesised voices in conversational AI systems.

Speculative Voicing Instagram (Abbas-Nazari, n.d)

Research Questions

In the thesis, I pose three questions. The first question addresses the development of a sonic speculative design methodology, and the sub-questions explore how this methodology can be applied to my research context. They are as follows:

1. How can thinking with and through sound develop a sonic speculative design methodology?
2. What does applying this methodology **reveal** about vocal profiling in the AI era?
3. How does applying this methodology **resist** vocal profiling in the AI era?

'Reveal' here means to expose the inadequacies and inner workings of vocal profiling in the AI era. 'Resist' means to inhibit vocal profiling to advocate for those who are marginalised by profiling in the AI era. Figure 5 summarises how my research questions are addressed through the thesis components to advance my original contributions to knowledge.

Research Questions	Methods	Methodology	Practice Projects	Analysis	Evaluation	Outputs
1. How can thinking with and through sound develop a sonic speculative design methodology?	Sonic Centric Speculative Design method developed in conjunction with appropriate participants	Combining theory from Speculative Design and Sonic Thinking to develop an intersectional approach.	Developed via series of workshops. Evolved and tested in case study projects.	Thick description observation of participants in workshop series and reflexive analysis.		Speculative Voicing Framework
2. What does applying this methodology reveal about vocal profiling in the AI era?	Vococentric Speculative Design, utilising 'Vacuum' metaphor in thesis writing.	Speculative Voicing Methodology	Case study projects employ Speculative Voicing Framework to create vocal imaginaries for human and synthesised voices.	Auto-ethnographic, synesthetic reflexive analysis of case study projects compared to current profiling practices.	IBM Workshop showed vocal imaginaries are obscured by deeply rooted relations between voice and profiling.	wav2face Google Colab. Speculative Voicing Workshop
3. How does applying this methodology resist vocal profiling in the AI era?	Vococentric Speculative Design, utilising concept of 'Polyphony' in thesis writing.	Speculative Voicing Methodology	Case study projects employ Speculative Voicing Framework to create vocal imaginaries for human and synthesised voices.	Thematic analysis of Speculative Voicing Methodology evidenced by participants of IBM Workshop	IBM Workshop showed strong potential for vocal imaginaries that resist profiling.	Voicing Beyond the Vacuum Max/MSP Patch. Speculative Voicing Workshop

Figure 5: PhD Summary table.

Aims

This research used speculative design to implement understandings of voice derived from sound and music practice, from the position of ‘sounding’ to question vocal profiling practices of/in conversational AI systems. Through investigating these fields and themes a sonic speculative design methodology emerged, which forms the basis of my original contribution to knowledge. The methodology was then applied to the original context to problematise vocal profiling through vocal imaginaries.

Original Contributions to Knowledge

My original contributions centre around developing a sonic speculative design methodology called Speculative Voicing. The methodology provides:

1. An intersectional new materialist position for speculative design by incorporating sonic thinking.
2. A means to investigate vocal profiling in conversational AI systems from a social and ecological standpoint, exploring being and identity.
3. A practice-based methodology of working with voice as a design material for vocal imaginaries.

Thesis Structure

My literature and practice review, in Chapter 1, explores current frameworks of understanding the sound and sounding in voices in conversational AI systems and how these are transcribed into vocal profiling, accompanied by a contemporary critique of these practices. Chapter 2 situates the voice in relation to my experiences as a singer, with two experiments that test and position my ideas. Chapter 3 discusses and unpacks a series of workshops that helped to develop my methodology. Chapter 4 describes how the workshop series is informed by and informs theory in developing an original experimental methodology. Chapter 5 tests the methodology that was developed and applied to human voices in conversational AI systems through the ‘Polyphonic Embodiment(s)’ practice project. Chapter 6

follows the same structure but is applied to synthesised voices in conversational AI systems through the ‘Acoustic Ecology of an AI System’ practice project. In Chapter 7, I document the evaluation of my work and methodology through a discussion of a workshop held with IBM employees. Chapter 8 concludes this PhD research with a discussion of the final workshop, a summary of my findings and outputs, details of the limitations of the research. I also include details of proposed future work, resulting from this research, towards informing more conscientious, responsible AI development.

This introduction summarises my PhD research and practice, which will now be developed in more detail throughout the chapters, as described. The next chapter, Chapter 1: Literature and Practice Review, discusses the key literature of the main themes in my PhD research, including the profiling of human and synthesised voices in conversational AI systems, the ethics of these practices, and an initial exploration of the voice as material and its polyphonic potential, using examples of existing practitioners.

Chapter 1: Literature and Practice Review

Introduction

This literature review explores how human and synthesised voices are currently recognised and understood in conversational AI systems through practices of profiling – both profiling humans from the sounding of their voices (Singh, 2019) and using profiling for the creation of synthesised voices and their sound design (West, Kraut & Chew, 2019). The ethical implications of profiling practices in AI are explored through their reliance on and reinforcement of normative expectations, drawing on Amaro (2019) and Birhane’s (2021) writing. This literature review reveals that profiling voices, in and by AI, is carried out within visual domains of understanding. I identify and utilise a metaphor of a vacuum and this functions as a point of reference to foreground AI vocal profiling’s neglect of the sonic materiality of voices. Neglecting the sonic further perpetuates discrimination because it aims to contain and constrict voices.

This literature and practice review has four sections. The first section summarises the gap in the literature and identifies where my original contribution to knowledge lies. Section two identifies key texts on the profiling of human and synthesised voices. Section three explores literature on AI ethics. Section four considers vocal materiality and polyphony, drawing on literature from sound, music, and voice studies. Throughout the text, I review related practice by other artists and musicians.

Literature Gap and Original Contribution

To date, literature that critiques and explores vocal profiling by AI systems is heavily contextualised and theorised via a position of *listening* – forensic listening (Abu Hamdan, 2018), ethnographic listening (Semel, 2021; 2020), machine listening and privacy (Lau, Zimmerman & Schaub, 2018), machine listening and racial bias (Koenecke et al., 2020), listening to misrecognition (Phan, 2022), biometric listening

(Kang, 2022), colonial listening (Vieira de Oliveira, 2021), and affective listening (Feldman, 2016). However, this thesis critiques the practice of vocal profiling via the position of the *sounding* of voices, primarily by understanding voice as material and having materiality. The sound of synthesised conversational AI voices has already been tackled from perspectives of gender bias (West, Kraut & Chew, 2019), anthropomorphism (Abercrombie et al., 2021), sociophonetics (Sutton et al., 2019) and the lack of diversity in synthesised voice design (Baird et al., 2017). Often, research of this nature focuses on biometrics, as this terminology aligns with the field of computer science and encompasses common examples such as voice, facial and iris recognition, and fingerprint scanning. While this research concerns biometrics, the terminology chosen was ‘profiling’, since this investigation is not exclusive to human voices. Revising the terminology allows for discussion that builds links with synthesised voices, as will be described further. Music practitioners have addressed conversational AI voices, often incorporating ideas of surveillance and privacy, by creating sound works or sound art, treating vocal sound material in a *musique concrète*¹² style. These include Kidel’s *Voice Recognition DoS Attack* (2018) and Jörg Piringer’s album *Darkvoice* (2019). This area is yet to be fully explored, and this thesis aims to bring ideas from music and sound performance but applied to vocal profiling using speculative design.

This investigation features the voice specifically as a point of departure and catalyst, giving it increased agency over and within conversational AI systems. I look to literature from sound-based practice to acquire richer, more nuanced notions of voice and vocal communication. In turn, this highlights a tension and disparity compared to existing understandings of vocal sounding in the fields of AI. A gap in the literature exists for a critique of profiling voices in conversational AI systems that engages theory from sound practice which is specific to discussions of voice. This

¹² *Musique concrète* is an experimental technique of using recorded sounds as raw material for musical composition. The technique was developed around 1948 by the French composer Pierre Schaeffer (Palombini, 1993).

research project focuses on vocal practice and the materiality of vocal sound, drawing on the work of Eidsheim (2015; 2019; Schlichter & Eidsheim, 2014). As I will discuss, vocal profiling presents the belief that it can present a very detailed picture of a person, and that voices can be contained within predefined categories of reasoning (Singh, 2019). This thesis argues that a voice is always multiple, as it is material and has materiality, hence the use of the term ‘polyphonic’. In addition, the voice in this research should be imagined as part of a choir or chorus, forming a constituent part of a whole.

Voice Profiling Practices in Conversational AI Systems

This section of text analyses how voices are recognised and interpreted within conversational AI systems. To undertake this, I review voice profiling frameworks which aim to identify humans by their voices in conversational AI systems and how the design and sonic aesthetics of synthesised AI voices in conversational AI systems follow these same frameworks. What becomes apparent is the reliance on notions of normativity within the defining of voices in this process. Ultimately, the validity and ethics of these frameworks are questioned and problematised through this PhD research and practice.

‘Voice’ and ‘speech’ are classified as two separate entities in AI-enabled recognition. As Singh describes, voice refers to sound produced in the human vocal tract. Speech is the signal produced by modulating voice into meaningful patterns (2019, p. 3). Here, an uncoupling takes place between speech (*logos*) and voice (*phone*)¹³ – a disembodied voice becomes multiple and is understood, computed, and analysed as such in conversational AI systems. Speech recognition concentrates on understanding linguistic¹⁴ elements of speech through natural language processing,

¹³ In his doctoral thesis, Lawrence Abu Hamdan writes extensively about this separation of *logos* and *phone* concerning speech recognition technology (Abu Hamdan, 2018).

¹⁴ As Mulder & Van Leeuwen (2019) note, ‘Linguists have rarely paid attention to the sounds of speech’.

text-to-speech (TTS) and speech-to-text (STT).¹⁵ This project is not concerned with the linguistic content of vocal communication in conversational AI (speech recognition). Instead, it focuses on and problematises the sound and sounding of voices (voice recognition) in conversational AI systems.

Interactive media artist Graham Pullin, writing with speech-language therapist Shannon Hennig in 2015, points out that sonic aspects of voice, such as tone of voice, is secondary in importance to linguistic content in synthesised voice production. Synthesised voices rely on STT and TTS processes that omit the voice's sonic qualities to focus on words.¹⁶ More recently, developments in AI-enabled machine listening have enabled a greater ability to derive information about people based on non-linguistic aspects of voice communication and have driven interest in this field of study. For example, Amazon has filed and been granted patents to determine users' 'real-time traits', including emotional and physical profiling. As well as analysing sonic voice data, these AI frameworks are trained to identify non-linguistic sounds (Jin & Wang, 2018).¹⁷ Although, for example, this patent uses a cough to signify illness sonically, it enables Amazon's Echo devices to attribute any sounded vocal cues to human bio-cultural features. The following section will discuss this in more detail, referencing writing by Singh (2019). Conversational AI systems use voice recognition and AI analysis to identify, distinguish and authenticate an individual by their voice.¹⁸ Alexa voice-enabled devices have built-in 'Alexa Voice ID', which will learn to recognise an individual's voice, call them by their name and create personalised experiences (Amazon, n.d). The extremes of voice recognition

¹⁵ Text-based natural language processing and its datasets have recently received notable critique, particularly Large Language Models (LLM) used in applications like ChatGTP (See: Bender et al., 2021). ChatGTP is not integrated into conversational AI systems such as Google Home. However, it can be added as a custom skill to Amazon's Alexa (Lombog, 2023), therefore it is anticipated that it will soon feature in all conversational AI-enabled devices.

¹⁶ This was also exhibited in the experiment 'Trying to Teach an AI to Sing' - See Chapter 2.

¹⁷ See also: Tech Transparency Project's review of Amazon's privacy policies and patent applications (TTP, 2021).

¹⁸ Also sometimes known as 'speaker recognition' (National Cyber Security Centre, 2019).

extend to voice profiling.

Profiling Humans from Their Voices

Profiling Humans from their Voices, by computer scientist and professor of language technology Rita Singh (2019), demonstrates contemporary voice profiling in conversational AI systems.¹⁹ Her book provides detailed accounts of how to infer 'bio-relevant facts' about people, their lives and their environment based on information embedded into human voices (p. viii).²⁰ AI correlates and cross-references data on sonic qualities of human voices to determine information as wide-ranging as 'Behavioural parameters: Dominance, leadership, public and private behaviour' and 'Physiological parameters: Age, hormone levels, heart rate, blood pressure' (p. 4). The spoken, linguistic content of voice data in voice profiling in conversational AI systems is often arbitrary.²¹ Singh describes profiling humans from their voices might be utilised for applications as broad as law enforcement, security, health services, social and commercial services and gaming and entertainment (pp. 366-369).

Singh recognises that, 'speech is as much a learned process as a natural, bio-mechanical one' (p. 99). Based on previous scientific studies of voice analysis, Singh proposes ways to mosaic historical research with contemporary machine learning and artificially intelligent systems (p. 2). In the quote below, Singh points towards the potential for patterning data to derive judgements that may have harmful implications for an individual:

¹⁹ John Baugh, in 2002, described how linguistic profiling resulted in housing being denied to people of colour in the US via telephone interactions, so this issue is not new or specific to AI. Additionally, Jennifer Lynn Stoeber, in their book *The Sonic Color Line* (2016), presents a cultural and political history of the racialised body and its relationship to emergent sound technologies.

²⁰ A recent high-profile example of voice profiling was start-up company Vocalis Health, which used vocal biomarkers to detect Covid-19 (See: Vocalis Health, n.d.).

²¹ For example, commercial company ClearSpeed creates voice AI analysis tools for companies who wish to screen potential job applicants to understand how trustworthy they are and identify potential candidates that are high risk. Potential employees are telephoned and asked a series of yes/no questions via an automated questionnaire. An assessment is then made using the sonic-derived data from these monosyllabic utterances (ClearSpeed, n.d.).

The web of relations deepens, and reinforces those between voice and the human face when myriad other links are considered. For example facial structure is related to a person's facial appearance.²² The relations of facial appearance to aggression,²³ and to race, and of aggression and race independently to voice, thereby connect face to voice. However we will not traverse such deep relations for now. Statistics, when used for prediction, must only be stretched so far! (Singh, 2019, p. 326).

Here, Singh identifies the potential for relations to be drawn between a human's voice and their face, race and potential for aggression using data science and statistical modelling for prediction purposes. The dominant understanding of the ability of the sound of voice to facilitate this has its roots in the field of phonetics. As Mulder & Van Leeuwen (2019) suggest, 'phoneticians saw speech sound as a symptom rather than a sign – an index of age, health, energy level, emotional state, and also of regional origin or ethnicity'. Singh notes that her above example would be going too far but does not recommend ways in which this could be avoided or specifically identify why this would not be desirable. Instead, this critical matter is bypassed by positing that race does not exist (Singh, 2019, pp. 99-100). While this position could offer a way to transcend and recognise the confines of categories and taxonomies, it is misaligned, as the goal is to enable extremely detailed profiling of an individual.

The code, applications and software to enable voice profiling by AI have been made widely available by companies such as ClearSpeed – 'voice analytics delivers a powerful vetting solution for fraud, security, and safety risk assessment screening' (ClearSpeed, n.d.), Vocalis Health – 'vocal biomarkers for personalised healthcare, screening and monitoring based on a patient's own voice' (Vocalis Health, n.d.),

²² Subtelny, J. D. (1959). A longitudinal study of soft tissue facial structures and their profile characteristics, defined in relation to underlying skeletal structures. *American Journal of Orthodontics*, 45(7), 481–507.

²³ Short, L. A., Mondloch, C. J., McCormick, C. M., Carré, J. M., Ma, R., Fu, G., et al. (2012). Detection of propensity for aggression based on facial structure irrespective of face race. *Evolution and Human Behavior*, 33(2), 121–129. Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, 20(10), 1194–1198.

IBM's Voice Surveillance tools (IBM, n.d.) and audeERING – 'Our technology can detect emotions and health information from the voice' (audeERING, 2021). A patent by Beyond Verbal goes as far as producing a 'Glossary of Tones', assigning musical notes to particular 'accepted emotional significance'. Examples include tone 'F' as 'Especially deep emotions such as love, hatred, sadness, joy, happiness' or tone 'B' as equating with 'Tones of command and leadership, ownership or sense of mission' (Levanon & Lossos, 2011).

Practices of profiling human voices in conversational AI systems are discussed and addressed in more detail in Chapter 5.

Profiling Practices in the Design of Synthesised Voices

I'd Blush If I Could (West, Kraut & Chew, 2019) is a report commissioned by UNESCO. The title of the publication borrows from the response voiced by Siri, a female-gendered voice assistant used by hundreds of millions of people, in reaction to a human-user saying, 'Hey Siri, you're a bitch'. The report details the gender biases designed into synthesised voices of conversational AI systems, of which the above example provides a potent portrayal of how these voices, projected as young women, perpetuate a negative impact.²⁴ Companies aim to design synthesised voices in conversational AI systems that sound as human as possible,²⁵ meanwhile, they claim that these voices are neither gendered nor human-like (Abercrombie et al., 2021).²⁶ Market research is conducted to understand and profile a company's consumers to create persona profiles for synthesised voices. The result aims to

²⁴ Golden Owens (2023) also writes about how Black voiced virtual assistants of conversational AI tune into the 'racialized sound of servitude in America'.

²⁵ For example, Google Duplex is an AI-enabled assistant with a synthesised voice created to sound as natural as possible and designed to emulate human communication, complete with prosody, pauses and punctuation. In the video of a demonstration during Google I/O, an annual developer conference held by Google in Mountain View, California, Google CEO Sundar Pichai has the AI assistant engage with a human to book a haircut (See: Mashable Deals, n.d.).

²⁶ Experiments found that despite the declarations from the products' designers, the analysis suggested that people interacting with these products tend to personify and gender the systems resulting from their design (Abercrombie et al., 2021).

design voices that consumers will be able to relate to or be sympathetic towards.²⁷ Designers create personified profiles for the illustration of synthesised voices. For example, James Giangola, a lead conversation designer and linguist for Google, describes how their conversational AI assistant was imagined during its design process:

A young woman from Colorado; the youngest daughter of a research librarian and physics professor who has a B.A in history from Northwestern, an elite research university in the United States; and as a child, won US\$100,000 on Jeopardy Kids Edition, a televised trivia game. She used to work as a personal assistant to a very popular late-night TV satirical pundit and enjoys kayaking (West, Kraut & Chew, 2019, p. 95).

Voice profiling, whether of human voices or in the creation of synthesised voices in conversational AI systems, presents a limited and naive understanding of voice. Sutton et al. (2019) note that advances in synthesised voices have concentrated on 'intelligibility and naturalness', which has aided usability and reliability. However, other properties and positions of voice have been neglected. The authors call for synthesised voices to be considered more critically and suggest incorporating 'made-up' accents for the voices by incorporating ideas from sociophonetics.²⁸ There have also been practice-based examples of experimentally synthesised voices (outside the field of music), including Q, 'the First Genderless Voice, created to end gender bias in AI assistants'²⁹ (Copenhagen Pride et al., n.d.). And [multi'vocal] is a synthesised voice which has been trained by using voice data from many people to create a diverse and collective voice (multi'vocal collective, 2021).

Practices of profiling for the creation of synthesised voices in conversational AI

²⁷ This is also common practice with human speakers in call centres. For example, banking firms locate their call centres in Scotland, as the accent aligns with economic common sense and trustworthiness (Aboutmatch, n.d.).

²⁸ Sociophonetics is the study of the social factors that influence the production and perception of speech (Sutton et al., 2019).

²⁹ The Q voice speaks between 145 Hz and 175 Hz, a range often classified as gender-ambiguous (Copenhagen Pride et al., n.d.).

systems are discussed and addressed in more detail as part of this thesis investigation in Chapter 6.

Profiling, AI & Ethics

Practices of profiling rely on normative and stereotypical representations of humans that cause real-world harm to people, especially those already marginalised within society. Benjamin (2019) accounts for the role that AI technology has in this: ‘bankers using financial technologies to prey on Black homeowners, law enforcement using surveillance technologies to control Black neighbourhoods, or politicians using legislative technologies to disenfranchise Black voters - which then get rolled out on an even wider scale’ (p. 32). Indeed, an expanding canon of literature examines, more specifically, aspects of AI recognition and ethics concerning racism (Hoffmann, 2018; Noble, 2018), genderqueer or trans people (Keyes, 2019; Costanza-Chock, 2018) and people experiencing poverty (Eubanks, 2018). This section will highlight two texts focusing on AI processes of categorisation and labelling and how these perpetuate normative expectations within AI systems applied to, in this case, conversational AI. This thesis project takes aspects of the critical discourse around these problems and aims to explore them through experimental sounding of voices. First, the discussion centres on Amaro’s (2019) ideas of ‘Black Technical Objects’. This text is highlighted in the literature review because it aligns with the aspirations of this research and practice to disrupt AI profiling schemas through a conscious, self-aware effort to uncouple from the lens of white, straight, patriarchal rationality while also finding new modes of belonging by reframing the lens of reasoning.³⁰ Then discussion then incorporates work by Birhane (2021) to argue that voice profiling in conversational AI systems currently exists within a metaphorical ‘vacuum’.

³⁰ This will be described further in Chapter 4, which sets out the intersectional methodology for this research.

Refusing Representation

The writing of Ramon Amaro, researcher in art and visual cultures of the Global South (2019), contributes to the vital discussion around racism and AI technologies.³¹ From the position of this thesis research, it would be hard to provide a new contribution specific to this field, or to claim to be explicitly decolonial. However, this thesis aims to support these themes, uphold the values and honour the ideas that are employed for this project.³² Rather than explicitly discussing race, this thesis uses Amaro's theory but broadens it to take an intersectional stance (Crenshaw, 1989). As will be explained, this approach is necessary because the problems surrounding AI recognition are intersectional: therefore the term 'normatively' is used to explore the current discourse around AI systems. Amaro's (2019) writing is contextualised with examples from artificially intelligent facial recognition systems. While this research is situated in the sonic modality of voice recognition, Amaro's text applies to any AI system that seeks to recognise aspects of human identity, personality or behaviour traits, and as such is still highly relevant. This is partly because AI voice recognition systems are also rooted in, and rely on, image-based analysis, in which sonic data is converted into waveform or spectrogram images.³³ However, more importantly, the same ambitions and aims underpin AI recognition and profiling frameworks.

³¹ Amaro's (2019) text is further expanded and developed in their book *The Black Technical Object: On Machine Learning and the Aspiration of Black Being* (2022).

³² When referencing decolonial texts or Black authors, I found strength and support in Audre Lorde's words: 'And where the words of women are crying to be heard, we must each of us recognize our responsibility to seek those words out, to read them and share them and examine them in their pertinence to our lives. That we not hide behind the mockeries of separations that have been imposed upon us and which so often we accept as our own. For instance, "I can't possibly teach Black women's writing -their experience is so different from mine." Yet how many years have you spent teaching Plato and Shakespeare and Proust? Or another, "She's a white woman and what could she possibly have to say to me?" Or, "She's a lesbian, what would my husband say, or my chairman?" Or again, "This woman writes of her sons and I have no children." And all the other endless ways in which we rob ourselves of ourselves and each other. We can learn to work and speak when we are afraid in the same way we have learned to work and speak when we are tired. For we have been socialized to respect fear more than our own needs for language and definition, and while we wait in silence for that final luxury of fearlessness, the weight of that silence will choke us. The fact that we are here and that I speak these words is an attempt to break that silence and bridge some of those differences between us, for it is not difference which immobilizes us, but silence. And there are so many silences to be broken' (Lorde, 1984 b, pp. 43-44).

³³ Discussed in more detail in Chapter 5.

This research also aligns with the position adopted by Amaro in which issues of racism and discrimination are rooted in the categorisation and quantification of people, which is a problem that pre-dates AI technology but which, unfortunately, computer science has subsumed at the heart of the frameworks that underpin AI recognition.³⁴ Amaro says Black people will always be interpreted through a white lens of understanding within AI recognition paradigms. From an intersectional position, the Black Technical Object can expand to signify any person who falls outside the bounded boxes of normative expectations of representation within systems rooted and embedded in white, patriarchal reasoning.

Amaro describes the Black Technical Object as ‘undetectable’. Here, he highlights how Black people struggle to be recognised and incorporated into machinic systems. One such example of this, in the context of voices in conversational AI systems, is the finding that all companies with significant automated speech recognition systems exhibited substantial racial disparities when trying to recognise Black voices compared to white speakers (Koencke et al., 2020). When viewed as a design problem, a solution to this failure could potentially be found with more research, testing and development. However, Amaro asks why Black people would want to be better included and integrated into systems that continually and disproportionately discriminate. This problem is particularly severe and increases exponentially in the context of artificially intelligent systems because, increasingly, they are used for decision-making in policing, crime detection and healthcare. Here, real-world harm is caused to the lives of Black people (Benjamin, 2019).

Borrowing from Moten and Harney,³⁵ Amaro suggests that the Black Technical Object is positioned to refuse representation, especially within a schema of universal

³⁴ Pasquinelli (2021) discusses more about how systems of classification became a ‘model of the mind’ that underpins artificial intelligence. However, this thesis does not address these ideas specifically.

³⁵ Harney, S., Moten, F. (2013) *The Undercommons: Fugitive Planning & Black Study*. *Minor Compositions*, 47–48. Moten, F (2008). The Case of Blackness, *Criticism* 50 (2) Spring, 177.

computation that limits self-determination. As Amaro discusses, the incompatibility of the Black Technical Object within AI recognition systems cannot dismiss existing racialisation. It may implicate it even further, but a feedback loop intersecting with these machinic systems could inform each other to ‘catalyse future affirmative iterations of self’. Here, the undetectable nature of a Black Technical Object causes a presupposed misalignment, misunderstanding or anomaly within a framework that could never understand the Black Technical Object in its totality.

This tension and friction between humans and aspirations of computability creates a new space for interpretation, understanding and being. Amaro (2019) asks, ‘How can Black Technical Objects generate new possibilities outside of phenotypical calculation, prototypical correlation, and the generalisation of category?’. Here, Amaro describes the frameworks of normative standards imposed by recognition systems and why Black Technical Objects fail to ‘fit’. Nevertheless, in his contemplation, he asks how Black Technical Objects might still find belonging and space to exist and grow beyond classifiers that contain and constrict being. Furthermore, he delineates and claims space for new possibilities and imaginaries to emerge through this formation. This endeavour of looking for emancipatory openings in existing constricted conditions also aligns with my interrogation of ways to resist vocal profiling.

Cathy Berberian (1925-1983) pioneered extended vocal technique, or experimental vocality – terms indicative of a diverse range of vocal utterances within vocal performance, much broader than lyrics-based song and singing. Berberian’s vocal practice inspired and paved the way for future generations of experimental vocal performers, many of whom are women, including Laurie Anderson, Joan La Barbara and Meredith Monk. As Vallee (2017) describes, in Laurie Anderson’s *O Superman* (1982), she used technology to ‘multiply her voice into many voices’ as a ‘critique of the voice image [...] in order to rupture the intersections of identity,

subjectivity, and body' (pp. 95–96). In addition, Karantonis and Verstraete (2014) highlight the cultural and social significance of Berberian's vocal aesthetics and style as a 'deconstruction of musical and spoken languages and visual markers of identity' within 'modes that privilege male-dominated concept[s] of authorship and a logocentric³⁶ way of understanding' and 'cultural meaning' (p. 5). Berberian used her vocal potential to disrupt the way she was expected to perform, be identified, and be seen within predefined frameworks of understanding. For example, in her composition *Stripsody* (1966), she can be heard voicing animals, including bees and dogs, tag lines from adverts, and objects, such as a grandfather clock. Berberian uncoupled herself from normative expectations by taking on a multitude of vocal aesthetics, and in her practice emancipated herself from the restrictive procedural elements of music performance. In this respect, we can compare Berberian and Anderson's vocal practice with an intersectional approach to Amaro's (2019) concept of Black Technical Objects. Both misaligned their vocal aesthetics to resist what was presumed and continually shifted sonic perception in order to prevent being categorised and contained by structures they found to be oppressive. Berberian also typifies a polyphonic approach in her sonically morphing into multiple beings and multiple others, that were potentially already contained within her.

Voicing in a Vacuum

Singh (2019) notes that human judgments of voices are limited by hearing ability, the brain's interpretation, and the physical and mental states of the listener (p. 10). She proclaims that ambiguity (a limiting factor for humans) would be eradicated through machine listening and AI computational processing, which provide faster, more accurate and larger scale comparisons (pp. 10-11). Singh is advocating for this areas of research, which she terms 'acoustic intelligence', to become a field of study in its own right (p. viii). In contrast, this thesis research champions ambiguity and an

³⁶ The author's use of the word 'logocentric' here borrows from Derrida's distinction between *logos* (word) and *phone* (sound): logocentrism in Western culture dominates, with words as signifying truth (See: Derrida, 1998).

understanding that people and their voices are ambiguous and multidimensional in order to unsettle the logic prescribed by profiling practices. As Birhane (2021) says in their paper ‘The Impossibility of Automating Ambiguity’, ‘In a worldview that aspires for certainty and predictability, the very idea of ambiguity, complexity, and multivalence—the essence of being, so far as there can be any—is not tolerated’. Therefore, without a tolerance for ambiguity how can the voice, as material and having materiality, ever be fully comprehended by an AI system?

Cognitive scientist Abeba Birhane (2021) writes that AI systems which fundamentally taxonomise and categorise humans are contained within a Cartesian and Newtonian worldview, in that they seek stability and predictability. However, she postulates that a post-Cartesian view of humans emphasises the indeterminable nature of a person and the entangled relationships between humans and others.³⁷ When computational AI analysis models are described as ‘working well’ this often equates to being ‘good at picking up historical patterns’, which compounds bias and confirms existing beliefs and predefined notions of normativity (Birhane, 2021). Therefore, by default, non-normative beings will always be classed as not fitting in, being wrong or anomalous.

Birhane³⁸ describes a Cartesian-Newtonian worldview as one based on objectivity – the assumption that observation, description and classification of the world can be done from a ‘view from nowhere’.³⁹ She also proposes that this is grounded in a white, straight ontology,⁴⁰ in which classification and prediction risk measure how closely people and their behaviours adhere to normative expectations

³⁷ I use a new materialist-based theory to explore these relationships in this research project. See methodology, Chapter 4 for more information.

³⁸ Birhane’s (2021) ideas are more extensively discussed and developed in their doctoral thesis: Birhane, A. (2022) *Automating Ambiguity: Challenges and Pitfalls of Artificial Intelligence*.

³⁹ Birhane borrows from T. Nagel (1989) *The view from nowhere*. Oxford University Press.

⁴⁰ Here, Birhane is referencing Sara Ahmed (2007). A phenomenology of whiteness. *Feminist Theory*, 8(2), 149-168. <https://doi.org/10.1177/1464700107078139>.

or socially and historically held stereotypes. She describes how data extraction, classification, and prediction processes used in AI systems, in turn, means that people are treated like objects. Birhane uses the term ‘objects’ to draw an analogy with derogatory ‘objectification’, but we can also understand it as people being assumed as static, without agency or active in the world.

Although not explicit in her writing, she describes the temporality and visual-based ontology of AI systems:

[...] relative stabilities and habitual patterns do not mean an individual person can be rendered fully knowable and predictable with precision. Any prediction of future behaviour based on past patterns is at best a statistical probability. We may, therefore, be able to predict a person’s general dynamics, under certain conditions, within certain context and time but precise prediction of a person’s specific behaviour and action, due to their nonlinear interactions and endless possibilities, are impossible’ (Birhane, 2021, p. 51).

In this respect, AI systems aim to predict (often future outcomes) but are, in fact, sequential or linear – built from pre-existing historical data but then held in that moment. Building on Birhane’s writing, I would like to extend the argument to suggest that AI recognition systems exist within a metaphorical vacuum.⁴¹ This will become a recurring metaphor throughout my thesis and this can be further explained with a short detour discussion about Le Grand K.

Le Grand K, or the international prototype kilogram (IPK) (Figure 6), is an object made from platinum alloy that sits in a vault on the outskirts of Paris; between 1889 and 2019 it defined the official standard of the kilogram weight. The object is contained within a protective double glass bell jar within an environmentally monitored and controlled locked safe, with the aim of keeping it in a vacuum-like,

⁴¹ Stoeve (2016) writes about listening, power and race. Borrowing from Du Bois (1903), Stoeve also uses a vacuum metaphor, describing it as ‘a barrier sound cannot cross. It silences black people within it, while enabling the white people outside to either ignore them or find amusement in their silent gestures of fury and frustration’ (p. 256).

airless environment. Despite these stringent preservative measures, for various reasons, both known and unknown, Le Grand K's mass can drift. For example, Le Grand K can gain mass because it has contact with air, and the atmosphere contaminates its surfaces. Any other interactions with the object, such as cleaning or handling, also caused it to gain or lose mass. When trying to maintain an international standard measurement, even the lightest fingerprint, with a mass of roughly 50 micrograms, has now caused deviation from this archetype (Wikipedia, 2021). Defining, standardising and measuring operations aim to preserve and create certainty and predictability (Birhane, 2021), through single, stable, knowable and identifiable occurrences. Whether this is the kilogram's weight or AI's recognition of voices, this highlights how aspirational acts of trying to define and measure phenomena require them to be extracted from the physical world in which they reside, removed from other life, matter and air. In other words, a desire to constrict voices within a knowable state requires containment to a vacuum, like Le Grand K aspired, rendering a voice devoid of its materiality.

Voices in conversational AI systems can only fundamentally exist when they can be transcribed in some way, such as TTS and STT (Singh, 2019, pp. 12-14). Human voice profiling in conversational AI systems, despite originating in the sonic domain, actually becomes an image recognition problem and procedure in AI systems, where the sound of voices becomes data via analysis of waveform or spectrogram images.⁴² An embodied voice transitions to disembodied sound to a solidified image. Voices in conversational AI systems are severed from the material world they inhabit, compressed into matrices of pixels.⁴³ These transcribed voices no longer exist as sound: they are forced to surrender their vibrational energy, materiality and active

⁴² Li & Mills (2019) chart the history of how spectrograms initially provided a visual record of features of individual voices to become a foundational model for understanding universal speech sounds. They describe how this transfer emerged from increasing demands for automated speech processing and aligned with a shift from the sound archive to the acoustic database.

⁴³ This process is discussed further in Chapter 5.

engagement in the physical world to the AI's vacuum. The sonic potential of voices diminishes and heightens the reliance on profiling practices.



Figure 6: International Prototype Kilogram (IPK). National Institute of Standards and Technology, US / Public Domain.

As hinted at by Birhane above, AI systems used in profiling practices can be metaphorically compared to photographs: one brief moment encapsulated by a boundaried box of the images' frame.⁴⁴ In this sense, we can consider AI recognition frameworks as existing in a vacuum, an airless container constricted by its walls. In stark contrast sits the modality of sound, including the sound of voices. Sonic material cannot be contained, captured or categorised,⁴⁵ which are the main functioning principles of AI, as alluded to by Birhane (2021). In contrast, as vibrationally contingent, sound knows no physical boundaries but is co-created and co-mediated with matter and the material it interacts with (Eidsheim, 2015).

This forms a mode of thinking to address and critique AI profiling practices of

⁴⁴ A helpful reference here is Carpenter and McLuhan's (1960, p. 67) ideas of visual vs. acoustic space - 'Auditory space has no point of favored focus. It's a sphere without fixed boundaries, space made by the thing itself, not space containing the thing. It is not pictorial space, boxed in, but dynamic, always in flux, creating its own dimensions moment by moment. It has no fixed boundaries; it is indifferent to background' (See also: Gow, 2001).

⁴⁵ As Eidsheim (2015) says, 'sound does not exist in a vacuum but is materially dependent'. This will be discussed further in the methodology, Chapter 4.

voices in conversational AI systems and will be a focal point of this thesis. A gap in the literature exists for theory from sound and voice practice to be applied to critical discussion of the sound and sounding of voices in conversational AI systems.

Voicing Materiality in Practice

This research is primarily a practice-led investigation of where the voice intersects with AI technology through the use and experience of my vocal practice as material to engage directly with its research subject matter. This section explores the specificity of this research as practice-led with further historical precedence and contextualisation initiated by experimental female vocal artists.

This research takes the position, as co-founder and former director of research at the AI Now Institute; Kate Crawford (2021) has documented, that artificial intelligence is neither artificial nor intelligent, as it aims to be portrayed. Instead, it should be understood as computationally enabled static modelling that is materially dependent on human labour (Taylor, 2018) and non-renewable resources (Crawford, 2021). A methodology derived from new materialist-based theory brings a material approach to this research to confront these issues, as will be discussed further in Chapter 4.

The study of the voice has been a focus in various disciplines such as film studies, linguistics, literature, performance, and anthropology. However, as Schlichter and Eidsheim (2014) note, a cohesive field of voice studies and a shared terminology has yet to emerge. In a position paper, Schlichter and Eidsheim describe that while discussion about the materiality of sound and its convergence with cultural, social and political realms have consolidated into the field of sound studies, the same is not true for discourse on voice. They observe that, historically, an understanding of the voice as an indication and marker of self in Western culture has turned it into an object – philosophical and theoretical enquiries of voice have under-appreciated its

materiality and rendered it 'symbolic'. They note that the material study into voice has resided within scientific-based disciplines like medicine, physiology and engineering. The authors, in contrast, offer a material understanding of voice where it plays an active role in 'human ecology', concurrently 'tied to the body and entwined with the external environment, the voice exists in a complex interaction with multiple physical and sociocultural formations'. I believe this understanding of voice and mode of thinking can play a decisive role in contributing to a discourse on vocal profiling in conversational AI systems.

Voicing Many Voices

I am drawn to Eidsheim's writing because the practice-led approach of this research is positioned from my experience as a singer and vocal artist. Eidsheim shares a background as a singer, and is a professor of musicology, so her writing is emphatically based upon her experience of working with her voice and traversing its possibilities, as this research also explores.

Eidsheim's (2019) book *The Race of Sound* explores the recognition of voices by their vocal timbre from a musicological perspective in relation to musical performance, practice and pedagogy. She dismantles the perception of race as 'an essential category' through listening to the sonics of voices. Particularly pertinent is her conceptualising of voices as part of 'thick events',⁴⁶ which follows from her previous book, *Sensing Sound: Singing and Listening as Vibrational Practice* (2015), which forms the basis of the methodology for this research. However, she suggests that voice is reduced to a stable and naturalised concept by asking, 'Who is this?' of an undisclosed voice. This tension is key to this research because Eidsheim's statement is also true of the way that voices are recognised in conversational AI systems, where their sound and sounding are discerned via predefined and pre-

⁴⁶ Eidsheim (2019) positions her writing in contrast to Cavarero (2005), who writes 'the human voice is 'a unique voice that signifies nothing but itself' (p. 5).

conditioned structures that intend to identify and typify people by their voices. She asserts that voice and vocal identity do not converge as a unified point of knowing or understanding. We can only comprehend voices ‘in their multidimensional, always unfolding process and practices, indeed in their multiplicities’ (Eidsheim, 2019, p. 3).

While Singh (2019) would consider this multiplicity a form of disguise or masking (pp. 15-17), Eidsheim’s sentiments are also echoed by contemporary experimental and extended vocal practitioner Jennifer Walshe. Walshe (2019) describes her body as a ‘staging area’ for all the things she has heard and all the places she has lived – ‘I don’t have a voice. I have many, many voices’. Music producer and performer Holly Herndon also explores creating voice multiples by training an AI, called Holly+, to mimic her voice, which she allows anyone to create music with.⁴⁷ These project examples are unified in using ideas of multiplicity and polyphony in creating and exploring music-making. This thesis research utilises these concepts to critique voice profiling in conversational AI systems while exploring the speculative potential for vocal imaginaries through practice.

Philosopher and cultural theorist Mladen Dolar describes the complex constitution of voice when he pronounces it as ‘a bodily missile which has detached itself from its source, emancipated itself, yet remains corporeal’ (2006, p. 73). In this respect, the voice is neither constrained, contained nor fixed but inhabits multiple states simultaneously. Voices in conversational AI settings, disembodied from their source, no longer have corporeal agency over how they are heard or comprehended. Atomised personal attributes and qualities are arranged in a formation that looks like, or centres on, a human individual, however, they could just as easily be

⁴⁷ Holly+ followed her doctoral investigations, which examined the ‘interplay between machine learning and the voice, and the implications of this technology for IP and vocal sovereignty’ (Herndon, 2021).

organised otherwise or taken independently (Behar, 2018).⁴⁸ The reconstruction and personification of disembodied voice signals in conversational AI systems take precedence over the actuality of the individual. We can regard the voice as always being both embodied – as emanating and ignited within the body, but also as disembodied – which animate and are animated by the materiality of the matter they interact with. This thesis further explores this embodied/disembodied tension by trying to mobilise all the multiple voices via their embodied origin and material potential. This multiplicity is vital to concentrate on when investigating conversational AI systems, because once the human voice leaves the body, it becomes an active participant and entangled in vast networked systems. By exploring this polyphonic potential of voice, this thesis aims to find more nuanced notions of voice to reveal and resist profiling practices.

Conclusion

Amaro and Birhane provide compelling and grounded arguments for scrutinising over-simplistic and normative expectations exploited by vocal profiling. Amaro describes that rather than turning against AI altogether, these systems can be folded into enquiries regarding identity and being, such as this PhD explores. Themes of intersectionality, materiality and polyphony led my research and practice to confront complexities of being and identity via Eidsheim's theories. Birhane gives rise to the metaphor of a vacuum, which I will now utilise. This metaphor serves as a reoccurring theme in this research and practice to move away from the currently prevailing ocularcentric ontology, as highlighted in this literature review, to foreground thinking and working with the sonic materiality of voice. Female experimental vocal practitioners help to support and evidence my ideas and practice, with additional contextualisation of my prior experience as a singer, discussed further in the following chapter. All these aspects combine to inform my

⁴⁸ Behar (2018) describes this as 'personalities without people' in the context of online micro-targeting marketing based on psychographics.

pursuit of revealing and resisting vocal profiling in the AI era.

Chapter 2: Contextualising the Voice as Sonic Material

Vocal Practice to Practice-Led Research

It is essential to discuss how this PhD, my practice and practice-led research, has been heavily driven and inspired by my background as a speculative designer and as a singer. In this chapter, I will identify how my prior experiences inform this PhD research and situate what it is about the voice that this research specifically concerns. The discussion of ‘Trying to Teach an AI to Sing’ and ‘The Voice...Sometimes Behaves So Strangely’ provides details of experiments that were initiated to understand better how AI recognition applications understand voice and the limitations in treating the voice in conversational AI systems as sonic material. These practice works also lead to contextualising and positioning this PhD research.

My singing experience has been gained through informal music education – singing in choirs. My vocal practice expertise falls into two categories: singing with choirs and, more recently, singing solo for several artists’ projects; these integrate to influence the practice and theory contained and highlighted within this research. The following sections describe these experiences and show that they inform the PhD research by the apprehension and appreciation of the voice as a material, polyphonic phenomenon and its conceptualisation in being part of a choir. The PhD work is different from these prior explorations as it seeks to formalise these vocal experiences through theory and practice, as and within a speculative methodology.⁴⁹ This is then explicitly applied to voices in conversational AI systems through the main practice case study projects of this thesis.⁵⁰

Choirs

I will first discuss the aspects of singing in choirs that shape this research work. A

⁴⁹ See Chapter 4 for more information.

⁵⁰ See Chapters 5 and 6.

choir consists of many individual singers who perform different 'voices'. These voices describe the different grouped singers' vocal parts, such as soprano, alto, tenor and bass, in a traditional choral formation based on vocal range. Each vocal part or voice follows the same section of notated music in a score. What I want to draw attention to is that in the context of a choir, a 'voice' is not singular. It cannot be attributed to one individual, body, or mouth. It is always multiple but acts in consensus as part of the composition. To sing in a choir is not as one voice in a crowd of many. It is to sing as one unanimous voice, with all the voices carefully blended, sonically unifying the individual into a cooperative, sonically reverberating cohesion. These personal experiences are echoed by Connor (2015), who describes this 'chorality' as a 'plural-singular' [...] 'collective voice-body'. A choir in performance can be considered an unfolding and becoming – as having many constituents but as being whole in and of itself. As an individual, as part of this event, a careful balancing of control takes place – disciplining the breath and body to contour the vocal output and surrender, allowing the force of the universality of the choir to carry one.

Vocal performance with a choir is a relational, event-based process mediated by listening. As a performer, it is necessary to incorporate listening to create an overall aesthetic of one unanimous voice in concord rather than highlighting specific voices or individuals. This process involves listening to each other and the space and environment in which the performance is taking place. Here, the environment is highly influential in defining how the voicing is enacted and heard. Acoustics can be considered a sonic representation of space, time, architecture, and materiality, animated by vocal sounding. However, acoustics also govern how specific voice qualities, such as diction and volume, are articulated. I want to highlight that I don't consider the voice to be singular and individual but rather as acting in a reciprocal, relational fashion with other factors such as space, time, material, matter and other bodies and voices. The (choral) voice is always multiple, always polyphonic in ways

where it is situated, relational and embodied because it is material and has materialism. These facets of a choir and choral singing became fundamental in developing this PhD's methodological framework, which aims to offer an alternative to AI's understanding of voice.⁵¹

During my time singing with Musarc,⁵² a performance platform investigating sound and space (Musarc, n.d.), one performance piece was very influential in my thinking about singing in relation to speech and bridging the two modes of vocality. Neil Luck's piece *Namesaying* (2013) guides performers with screenshots of mouths captured in vocalising speech sounds. Performers are asked to view the images and vocalise the sound they think the image represents. This performative experience catalysed my consideration of speech from being quite reserved and conservative to having real vocal sonic potential and musicality. A distinct gap exists if we compare the sounding of speech to the broad breadth of sound produced by a singer. A creative, intriguing gap.

Female performers pioneered extended or experimental vocal techniques, incorporating speech sounds and other unsung vocal noises. However, these considerations remain within the realms of music and performance. Here speech, or voice, is sonic (not exclusively linguistic). A brilliant example is Elaine Mitchener's performance of Christian Marclay's graphic score *No!* (2020). Watching Mitchener perform⁵³ this piece at the London Contemporary Music Festival (2022), I observed her morphing her mouth, face, and entire body to perform what seemed like every

⁵¹ See Chapter 4 for more information; the chapters which follow test and analyse the methodology, as a key contribution of this PhD research.

⁵² Thanks also to Musarc and their creative director, Joseph Kohlmaier, who has been a long-standing supporter of my work, especially since the Field Studies Symposium, where I presented some very early thoughts about this PhD research before applying for study. Writing at the time of the symposium, I describe how my work 'employs voice as a medium, to be shaped, sculpted and moulded to investigate "where speech meets sound", blurring these boundaries and exploiting vocal potential to devise sonic fictions - stories about alternate arrangements for society via design, technology and politics.' (Field Studies, 2017)

⁵³ Documentation of Elaine Mitchener performing of Christian Marclay's graphic score, *No!* is also available online (See: Fraenkel Gallery, 2021).

possible way of voicing 'No'. Each iteration of 'No' had its own meaning, motivation and poignance, animating one simple syllable into many possibilities. This thesis research argues that female experimental vocalists set precedents for, and provide evidence of, the voice as material and polyphonic as they exhibit in their practice. They repeatedly demonstrate that voices can and do exist beyond AI's current comprehension of vocal sounding.

I first started to investigate and grapple with the themes discussed above during my MA studies in Design Interactions at the Royal College of Art, primarily through my final major project, *Across the Sonic Border (Variations on 50Hz)* (Abbas-Nazari, 2014). During the MA course, I explored voice as a medium within speculative design practice, particularly concerning emerging technology. However, I can trace my practice exploration of voice in conjunction technology to my undergraduate studies, around 2009.⁵⁴ I note that these observations are only available to me because of the privilege I hold and recognise in having the opportunity to study the voice and to be able to sing. As a mixed-race person, half-Iranian, half-British, I am always aware that this practice would be arduous to conduct in Iran today, where how, where and to whom women sing is a highly restricted and constrained endeavour. The tension between wanting to creatively explore the full range of human vocal potential in voice communication whilst understanding that the sound and sounding of voice(s) can be highly contentious within different settings is also a recurring theme in my work. For this research, this tension is present and explored within voice profiling practices in conversational AI systems.

Solo Vocal practice

Since 2016, I have performed solo on several artists' projects, and these propositions have explored and enhanced my experimental and extended vocal

⁵⁴ The Speculative Voicing website (Abbas-Nazari, 2022) documents my voice-related practice since this date.

practice expertise. For example, I have been asked to imagine and vocalise half-human, half-animal sounds for artist Fani Parali, which are then intricately lip-synced by other performers (See: Parali, n.d.). The process of working with Parali is intriguing. I remain removed from the work, and my contribution is only revealed as I watch the assembled performance as an audience member. I hear my disembodied voice being absorbed and exuded by a quasi-mythical being, far removed from my sense of identity. Whilst this experience is pleasurable for me, it could also be compared to the lack of autonomy people may feel when an AI reflects misinterpreted conceptions of their identity via their voice alone.

I have also produced imagined vocal sounds of humans first finding consciousness and the origin of language for the artist Marguerite Humeau (See: artviewer, n.d.) and the sound of black holes, nebulae and other cosmic matter for artist the Nestor Pestana (See: Pestana, n.d.). This process involves an orientation towards a disembodied embodiment – trying to encapsulate time scales, bodies, realities and matter wholly removed from my being. Placing my imagination in these spaces, I morph my vocal apparatus, mouth, face, core muscles, and bodily architecture. The endeavour is to become a vessel to shape air and resonate my materiality as these heterogeneous bodies would. It is a process of sounding and concurrent listening for feedback to gauge the attempt made – imagining and then embarking on trying to illustrate that imaginary sonically.

Listening functions as a mechanism to both situate the sound and as a response to reflect on the catalysing sonic action. As described above, the sounded voice shape-shifts through many factors, such as architecture and acoustics. Upon returning to the body via the ear (primarily, but not entirely), a fuller appreciation of the voice is heard in its multiplicity and entirety. I align my understanding and experience of voice with artist the Mikhail Karikis (1997), who considers '[t]he voice as a sculptural material' which can be 'stretched' and 'manipulated'. The experiences

I describe here will be further contextualised and deepened through the practice projects in this thesis and supported by existing literature in developing my sonic speculative design methodology.

Trying to Teach an AI to Sing

The experiment ‘Trying to teach an AI to Sing’ involved creating a synthetic clone of my voice using an AI application called Lyrebird ([Item 2](#); [Item 3](#)), an online application now acquired by Descript (n.d). The training involved speaking around 40 short sentences aloud and then one hour of analysis and processing by the AI. Once completed, I could use my cloned voice to create text-to-speech synthesis. In ‘Trying to Teach an AI to Sing’, I attempted two approaches. One failed, and the other produced some interesting but limited results. I first tried to teach the AI to sing during its learning process by wildly varying the pitch and vocal emphasis of speech sounds of the sentences I was prompted to speak. This method soon failed, as any marked variation in the voice input was detected, and consequently, I was repeatedly asked to ‘try again’. The second approach was to intervene in the text-to-speech output. In this test, I typed a mixture of nonsense words, isolated consonants, and extended sequences of vowels. Here, some similarity to singing could be heard ([Item 4](#); [Item 5](#)). Perhaps because there was variation in the tonal range of the voice, and the voice sounded quite different from how I usually speak, I associated it with singing.

The experiment highlighted that while the AI attributed and paired voiced sounds to particular phonemes of the language spoken, the application did not learn my voice’s overall sound and sonic qualities. The AI focused on producing speech instead of learning and mimicking a voice. While the AI may have cloned my speech and contained some element of the sound of my voice, it did not acquire the so-called ‘grain’ of my voice (Barthes, 1977) or my voice as a whole. The AI application would make choices of intonation between syllables and words that were unexpected and

unfamiliar in my use of voice. It was noticeably limited compared to my full vocal range and capabilities. For example, it would be particularly inadequate in producing vocal gestures, such as a scream, cry, or gasp.

The first synthesised voice to sing was produced by an IBM 704 computer at Bell Labs, programmed to sing the song *Daisy Bell* in 1961 (Radovic, 2008). A more contemporary example of sung voice synthesis is Holly+, an AI-trained synthesised voice clone of singer Holly Herndon, created in 2021. However, to create this singing voice clone, Herndon had to train the AI with her sung voice, creating a separate AI and training model for her spoken voice (Holly+, n.d.). She would train the Holly+ 'Speaking Model' by mapping her spoken voice to units of speech such as phonemes, and the 'Raw Singing Model' would be trained by mapping her sung voice to musical units of sound such as tones and semitones. Through my observations of projects such as Herndon's and my experiments with Lyrebird, I deduced that there is a voice/speech distinction in AI; this is also reinforced by the relevant literature.

In the experiment of 'Trying to Teach an AI to Sing', my research and practice questioned the sounded qualities of speech and voice in conversational AI systems. Questions arise about where voiced speech becomes vocal sound, and where vocal sound becomes voiced speech. Or at what point does vocal sound lose or gain linguistic or semantic meaning? For example, I comprehend a voice speaking in the Iranian language of Farsi, a language I do not understand but recognise through its sounding, as song. Nevertheless, another question might be: what is the difference between the voice as spoken or sung? These are not questions I aim to answer definitively. Instead, they are proposed to advance towards a hazy space between these two states of voicing and understanding, where the singer and speaker

converge and collapse.⁵⁵ Here, speech sounds can be utilised for their musicality, which supposes they have creative malleability beyond their role as carriers of purely linguistic information. In ‘Trying to Teach an AI to Sing’ the ‘rules’ of language and linguistics programmed into Lyrebird restricted the sonic possibilities of voice I know to be possible via my previous experiences as a vocalist. I intend to explore this further in this investigation,⁵⁶ exploring the voice as sonic (not linguistic), and the polyphony this affords.

The Voice...Sometimes Behaves So Strangely

Psychologist Diana Deutsch demonstrates how repeating a recorded segment of vocalised speech becomes song in her ‘speech to song illusion’ experiment (Deutsch, 1995). When we listen, there becomes a point at which the sonic effect overrides our linguistic comprehension of the words as we trace the melody in the phrase spoken. Artist and audio investigator Lawrence Abu Hamden (2018) describes the practice of ‘forensic listening’, in which spoken words become sonic specimens, allowing for analysis of vocal sound in the assessment of a person’s character, behaviour and identity, saying ‘not only our words (*logos*) but our voices (*phone*) can be made to testify’.⁵⁷ In Abu Hamden’s doctoral thesis, he recounts an interview with a forensic linguist, who describes: ‘Last week, a colleague and I spent three working days listening to one word from a police interview tape’ (p. 57). Here, the repetition of a recorded and captured voice enables a vocal profiling assessment. This is equally true of voice identification and analysis in conversational AI systems. As Edward B. Kang (2022), professor of critical digital studies notes, this industry relies on understanding the voice as a ‘fixed, extractable, and measurable “sound object”

⁵⁵ These ideas are more definitively explored by Cummins (2020), who hypothesises a speech-music continuum.

⁵⁶ In my research, I struggled to navigate language and linguistics, combined with my intentions, for a long time. I attempted to engage with these fields via Berardi’s (2018) idea of poetry as an ‘excess’ of information that cannot be reduced to pure data. However, as it became more apparent that voice and speech are separate in conversational AI, it made sense to concentrate my focus on ‘voice’, as my expertise through singing dictated.

⁵⁷ Abu Hamden’s work concerns the political effects of listening and its impact on human rights and law.

located within the body'. This ideology maintains that people only have one voice and that someone's voice can be used to measure the person. Abu Hamden's example describes how this is possible with an extracted, recorded and captured voice. In AI-related examples, machine listening replaces the labour of the human forensic listener.⁵⁸

In reality, the sound and sounding of an un-extracted or embodied voice is never fixed. Recorded on the hour, over eleven hours of a day, I tried to replicate the voicing of the phrase 'sometimes behaves so strangely' for 'The Voice...Sometimes Behaves So Strangely' ([Item 6](#)).⁵⁹ This demonstrated a failure to maintain the aspired consistency of the 'speech to song' experiment, in which Deutsch's repetition of the phrase was edited using technology. In the logic of AI's vocal profiling, afforded by machine listening, the desire is to zoom in and measure each utterance contained within the phrase, in order to read and fully comprehend the articulated sonics of my voice. Perhaps I was heard to say the words louder and faster due to a caffeine intake. Alternatively, perhaps my voice deepened in pitch, indicating that I was tired. However, from the perspective of sounding (as opposed to listening), the embodied, sounded voice always evades capture by the listener, whether human or machine, through its ever-dynamic materiality. I believe voices can gain increased authority and autonomy within conversational AI systems from the position of sounding.

As a trained singer interested in extended and experimental vocal technique, my understanding of voice aligns with singer and musicologist Nina Sun Eidsheim. Writing in 2019, she describes how 'a specific voice's sonic potentiality [...] [in] its execution can exceed imagination' (p. 7) and discusses voices as having 'an infinity

⁵⁸ As described in my literature review, there is a growing body of literature about how AI can be applied to analysing voice to determine wide-ranging factors about an individual.

⁵⁹ This phrase is adapted from Deutsch's (1995) experiment in which she says, in full, 'The sounds as they appear to you are not only different from those that are really present, but they sometimes behave so strangely as to seem quite impossible'. 'sometimes behave so strangely' is the phrase she used to exemplify her illusion.

of unrealised manifestations' (p. 8). In an understanding of the voice as material and as having materiality, it is no longer singular, stable or consistent, but it can morph, shape-shift and have polyphonic potential. Here, voice can simultaneously occupy the space of speech and song, becoming a sonic material to be shaped. This research focuses on 'sounding', and on what this has to offer in the field of speculative design to examine vocal profiling practices. Listening is still present, but as a singer would listen. It acts as a feedback function to hear voices in their entirety through their materiality in conjunction and co-creation with other matter.

Conclusion

This reflection-on-action (Schön, 2016) of my prior experiences and two experiments help to contextualise my research and position my understanding of voice. To summarise, I will consider the (spoken) voice as sonic material with polyphonic potential, especially from the position of sounding, as opposed to listening. Understanding some of the current limitations and constrictions of voicing from these two experiments, I will now move to focus on the sonic, describing how I took a sound and voice-led approach in my practice-led research. In order to do this effectively, my research did not employ AI or machine learning as a primary medium in creating the works.⁶⁰

This chapter situates the three research questions of this thesis.⁶¹ These identify the aim of amalgamating and elaborating on my experiences of vocal practice and of studying speculative design to generate the original contribution of a sonic speculative design methodology. It will be developed further through existing theory and practice from sound and music.⁶² I will apply this methodology to explore vocal sounding and profiling in conversational AI systems, aiming to find

⁶⁰ Discussed in more detail in chapter 4.

⁶¹ See the Introduction Chapter for more details.

⁶² Described in the following two chapters.

ways to reveal and resist vocal profiling in the AI era.⁶³

⁶³ Predominantly explored in Chapters 5, 6 and 7.

Chapter 3: Participatory Workshops - Developing a Sonic Speculative Design Methodology

Introduction

This section details six workshops I devised and facilitated, which led to the development of an experimental sonic speculative design methodology. Four workshops, titled 'Speculative Listening' (I–IV), all followed the same structure and process (see below). Specific details regarding location, premise, participants and participant outcomes, along with reflections and analysis relating to each of the workshops, is accounted for under their particular headings. Two further workshops, titled 'Multiphonic Connections' and 'Giving Voice to Synthetic Sonics', are also detailed: these, through their analysis, enabled the validation of my findings concerning the development of an original experimental methodology. The workshops are documented in chronological order to show the iterative nature of the practice-led work, and the Final Reflections and Conclusion describes how this work contributes to the next steps in the PhD research. The main principle of the workshops looked at how sound and sounding can be used as a guiding force to imagine novel and speculative sonic interactions. This chapter aims to address the first question in my PhD enquiry: 1. How can thinking with and through sound develop a sonic speculative design methodology?

The next chapter describes the PhD research and practice relating to the way that these workshops inform and are informed by existing theory in the development of the methodology. The methodology was later tested in scenarios specific to the sound and sounding of voices in conversational AI systems (see Chapters 5 & 6) and evaluated as part of an additional workshop with employees from IBM (see Chapter 7).

Speculative Listening Workshop Structure and Process

During all four ‘Speculative Listening’ workshops (I-IV), I asked participants to imagine what sounds they would like to listen to that we currently cannot with our human ears alone. This provocation was catalysed by providing participants with a soundtrack of ‘inaudible audio’ ([Item 7](#)) – sounds we can only hear with the help of technology. The 9.5-minute soundtrack consists of 18 different, short, found sound clips, including sounds from space, deep underwater sounds, the sound of bats and that of corn growing.⁶⁴ An AI replica of my voice narrated this playlist.

Participants were also shown images of the artist Nick Cave’s *Soundsuits*, a series of body-worn sculptures made from diverse readily available materials, to inspire them to augment and incorporate their bodies with their designed devices. This also, subtly guided them to consider the politics of sounding and listening in relation to identity during the workshop. Cave’s collection of works, which he has been making since 1992, provides anonymity by concealing any visible markers of a person’s race, gender and class.⁶⁵ In an interview for *The Washington Post* Cave explains: ‘I built this sort of suit of armor, and by putting it on, I realized that I could make a sound from moving in it [...] It made me think of ideas around protest, and how we should be a voice and speak louder’ (Hoo, 2012).

In the workshop, I guided participants to contemplate the growing capabilities of AI-enabled machine listening and its role in privacy and surveillance. We discussed how technological developments might be redirected to listen in other ways. We asked: ‘How do we want to be heard by technology?’ ‘How can technology allow us to be heard and listen in different ways?’ ‘How does what we hear shape the way we understand the world and each other?’. Participants addressed these questions by sonically thinking through making, using simple materials, such as card, foil,

⁶⁴ For full track listing with references, see Appendix A.

⁶⁵ For documentation of Cave’s *Soundsuits*, see: Madeleine (2014)

balloons, and string (Figure 7), to construct 'listening devices' or 'body microphones' to describe their ideas. Participants were invited at the end of every workshop to express their ideas to the group. Their 3D-made objects functioned as prompts to help describe imagined, speculative sonic interactions.



Figure 7: Materials for Speculative Listening workshops. Lisa Marie Bengtsson.

These workshops engaged with young people aged 3 - 25 and those working with young people (Speculative Listening II). Research has shown that this age group have more fluid conceptions of gender and identity (Twenge, 2023), are more ethnically diverse than any previous generation (Bakhtiari, 2022)⁶⁶ and are potent in activist movements for social change (Carnegie, 2022). It is appropriate to work with these people to develop my original methodology because it is vital to generate new ways of working that reflect changes in attitudes and society, to address issues of categorisation and marginalisation that are prompted by AI.

⁶⁶ This reference considers the population of the USA, however census data from the UK mirrors these findings (See: gov.uk, 2023).

Speculative Listening I

Location: Barbican Centre, Life Rewired Hub

Premise: Invited by artist and play worker duo India Harvey & Lisa Marie Bengtsson, who run Squish Space at the Barbican Centre. Squish Space offers multi-sensory play for children under five. The workshop was part of a day of activities called 'Future Landscapes of Child Culture' as part of the temporary Life Rewired Hub (Figure 8) within the Barbican Centre. The Life Rewired Hub programme ran from February to December 2019, offering 'talks, performances and residencies [...] invite[ing] audiences to engage with the dizzying impact of technological and scientific change on what it means to be human today' (Barbican, 2019).



Figure 8: Life Rewired Hub, Barbican Centre, sectioned off with semi-transparent curtain.

Participants: ~15 children aged three to seven, accompanied by their parents. Participants were self-selecting: they either found out about the workshop via the Barbican website or may have happened upon the workshop because they were visiting the Barbican that day, 27 April 2019.

Participant Outcomes: Participant ideas were verbally expressed in a minimal capacity, owing to participants age and limited communication skills. Most made 3D objects without fully describing what they were (e.g. Figure 9). Other outcomes included a device to hear spiders,⁶⁷ a fart-amplifying machine and a device for communicating secret messages.



Figure 9: Speculative Listening device made during Barbican workshop. Lisa Marie Bengtsson.

Reflections and Analysis:

For this first ‘Speculative Listening’ workshop (and the following one), I was fortunate to be offered the use of 25 sets of silent disco headphones, which were suggested and organised by Harvey and Bengtsson.⁶⁸ This created a magical, intimate yet communal experience, since everyone wearing the headphones hears

⁶⁷ This and many other participant outcomes could be framed as voices, e.g. the ‘voice’ of a spider. While this approach could be beneficial in acknowledging the entangled, more-than-human leaning of this research, for clarity in the PhD project, ‘voice’ is only used to denote those that are human and/or synthesised.

⁶⁸ These headphones work by receiving an audio signal broadcast by a transmitter, like FM radio.

the same thing simultaneously. We used the headphones to listen to the specially curated soundtrack of 'inaudible audio' ([Item 7](#)), as well as communicating the instructions of the workshop by speaking into a microphone for everyone to hear (Figure 10). The soundtrack played on repeat throughout the workshop unless someone spoke into the microphone. Participants were free to wear or remove their headphones at any time.



Figure 10: Using the silent disco headphones and microphone during Barbican Speculative Listening workshop. Lisa Marie Bengtsson.



Figure 11: View of workshop taking place at Barbican, from above. Lisa Marie Bengtsson.

Even though I am familiar with Squish Space and Harvey and Bengtsson's work, I naïvely neglected to consider how young children would be the main participants of the workshop! As a result, I had to quickly modify my verbal instructions, using simpler language to describe the workshop activity. This included describing the objects they were about to make as 'body microphones' and 'listening devices'. This allowed me to address how participant ideas may include sound emanating from the body and sounds being received by the body, which was communicated in future workshops. To describe how I wanted the children (and their parents) to consider the multi-sensory nature of listening and the whole body in their design ideas, I gave the example of how elephants 'hear' with their feet (Yollin, 2007), which intrigued the children. Some children understood that their devices should make or create the sound rather than describe an imagined sonic experience. I realised this was an important distinction I needed to make for future workshops. The workshop was fantastically messy and chaotic (Figure 11). Many of the children made things together with their guardians. Other children played with the materials or chose instead to draw while adults spoke to each other. To conclude the workshop, the children (no adults) came together to show each other what they had made (Figure

12).



Figure 12: Children taking it in turns to present their devices. Lisa Marie Bengtsson.

Speculative Listening II

Location: Tate Modern

Premise: Invited by India Harvey and Yemi Awosile as part of Tate Modern's Summer School 2019, which ran for a week and appealed to people to 'think about teaching in an expanded sense' and 'contribute to an evolving, experiential and participatory conversation around new approaches to teaching and learning' (Tate Modern, 2019). The 'Speculative Listening' workshop was programmed by Harvey and Awosile as part of activities devised to encourage participants to think about neurodiversity and (multi-)sensory diversity in relation to learning, teaching and participatory ways of working in education and/or cultural spaces.

Participants: ~16 participants, a mix of art teachers and museum educators. About a third were based in international schools and had travelled to London to

participate (Thomson, 2019). Participants were self-selecting in having signed up for the Tate Modern's Summer School 2019 programme.⁶⁹

Participant Outcomes: devices to hear trees communicating, the sound of conception – when sperm meets the egg, for sex education purposes (Figure 13), listening to ghosts/voices of the dead, listening to deep underground sounds via the hip bones (Figure 14), communicating or understanding subconscious thoughts and workings of the brain, and communicating/making audible all the 'extra', non-verbal information we transmit while speaking. Some participants chose not to share their ideas.

⁶⁹ A fee was payable to Tate Modern for this course, as set by the institution. However, bursaries were available. My workshop was part of a week-long programme of events for the Summer School.



Figure 13: A Speculative Listening device to hear the sound of conception - when a sperm meets the egg for sex education purposes. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys.



Figure 14: A Speculative Listening device listening to deep underground sounds via the hip bones. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys.

Reflections and Analysis

For this workshop, I was again supported by Tate to hire silent disco headphones. In the introductory discussion about the workshop, we collectively talked about the conditions that limit humans' capacity to hear certain sounds. These included sounds that are too far away, too high or low in pitch, that occur too slowly, are too quiet, or might be hidden (e.g. under/inside something). We noted these on A4 paper for everyone to consider (Figure 15) while listening to the 'inaudible audio' soundtrack and exploring their ideas with the materials provided. We also talked about different styles/types of listening, including deep listening (Oliveros, 2005), reduced listening (Schaeffer, 2017) and active listening (Rogers & Farson, 2021), which I invited them to try out during the workshop. This promoted contemplation by the group on different ways to interpret, understand or respond to sound and sounding, especially with the added context of neurodiversity and (multi-)sensory diversity in education and learning. I then articulated initial ideas of how

speculative listening, as a concept, might be a way to consider imagined sound and sounding to speculate on novel sonic experiences. I was surprised and delighted that participant ideas were very creative, insightful, and nuanced, and this was achieved with little or no additional prompting or guidance (other than described).

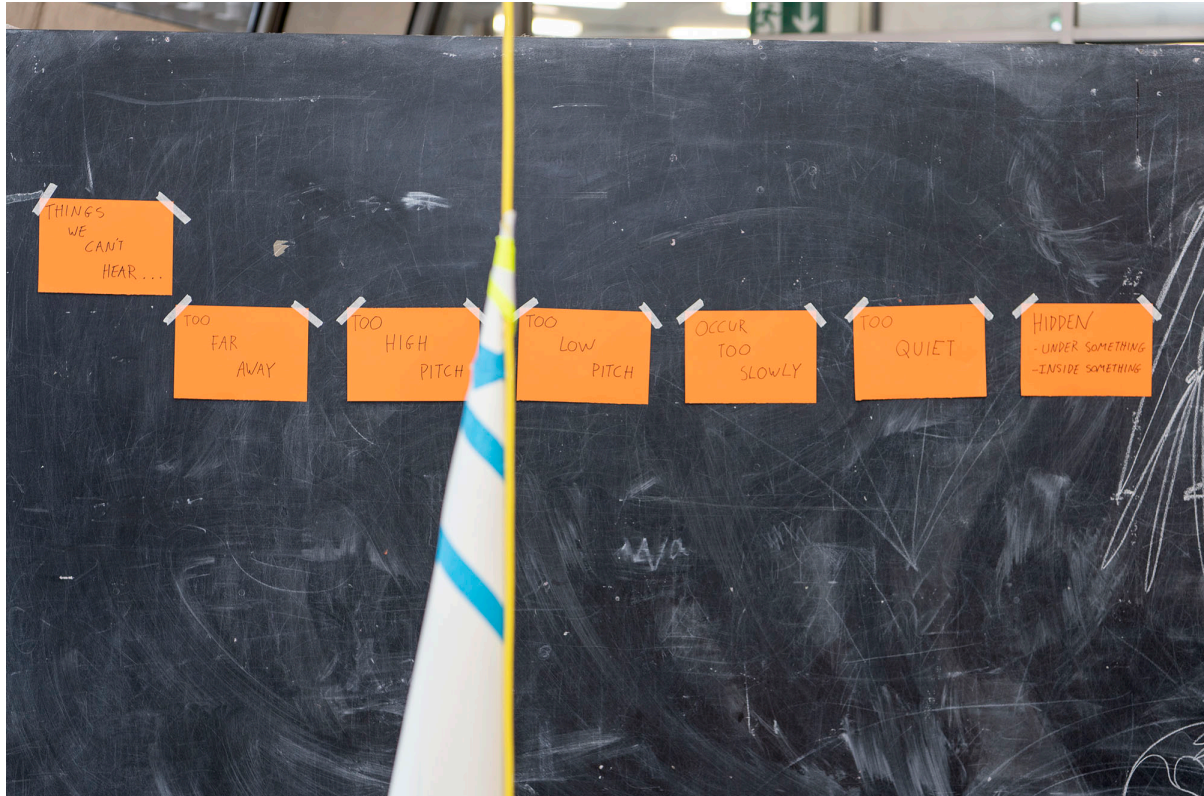


Figure 15: Possible conditions that limit humans' capacity to hear certain sounds, Speculative Listening II workshop. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys.

Speculative Listening III

Location: Leeds University, Art & Design School

Premise: Invited by Deborah Gardner, lecturer in undergraduate art and design, Leeds University

Participants: 5 undergraduate design students. Participants were self-selecting via an email to the department from Deborah Gardner, who had invited me.

Participant Outcomes: a device for individuals to listen to their menstrual cycle to

better understand mood changes and communicate this to others (Figure 16), sonification of chaos and personal decision-making activated by the breath (Figure 17), being able to sonically describe anxiety, emanating from the stomach (Figure 18), listening to rocks and fossils to get a sense of deep time, hearing emotions and being able to comprehend them beyond specific categories – happy or sad, for example.



Figure 16: A Speculative Listening device to listen to the menstrual cycle to better understand changes in mood and communicate this to others.



Figure 17: A Speculative Listening device for the sonification of chaos and personal decision making activated by the breath.



Figure 18: A Speculative Listening device to sonically describe anxiety (emanating from the stomach).

Reflections and Analysis

The use of silent disco headphones was not possible in this instance, due to lack of funding. However, I aimed to recreate the experience by uploading the ‘inaudible audio’ soundtrack to the Soundcloud website. I asked participants to come prepared with headphones and an internet-enabled device to access the audio via a link provided. This worked well as an alternative, enabling participants to collectively listen to the sound clips. In this workshop, we did not discuss different types or styles of listening apart from mentioning how listening is an activity that can incorporate the whole body, not just the ears (Ihde, 2007). In addition to my prevision of materials, in this workshop and Speculative Listening IV, participants were requested to bring simple (clean) materials to share, noting that they could probably find these around the house and/or in their recycling bin. Again, I was compelled by the poetic nature of participant responses that showed significant curiosity and sensitivity towards the lived experience and ecological thinking from a

more-than-human and entangled perspective.

Multiphonic Connections⁷⁰ (Reworked Speculative Listening Workshop as Interactive Telephone Experience)

Location: Interactive Telephone Experience enabled by Zoom.us/Online

Premise: Invited by students from the MA Curating Contemporary Art Programme Graduate Projects 2020, Royal College of Art, as part of 'Empathy Loading' in partnership with Furtherfield Gallery's 'Love Machines' summer programme – an online project exploring empathetic relationships between humans and networked non-humans (Empathy Loading, 2020).

Participants: Available between 18th and 21st June 2020; the interactive telephone experience had 64 callers, and ten callers left voicemail messages. Two people contributed to the work via social media instead of leaving a voicemail. The work/telephone number was advertised via poster images (Figures 19 & 20) on social media and the Empathy Loading website (See: Empathy Loading, 2020). Abbreviated quotes from voicemail messages received were used in additional social media posts (Figures 21 & 22) to encourage further participation.

⁷⁰ The title of this project borrows from composer Maryanne Amacher's (1938-2009) fictional company 'Supreme Connections' in her unrealised media opera *Intelligent Life* (See also: Cimini, 2019).

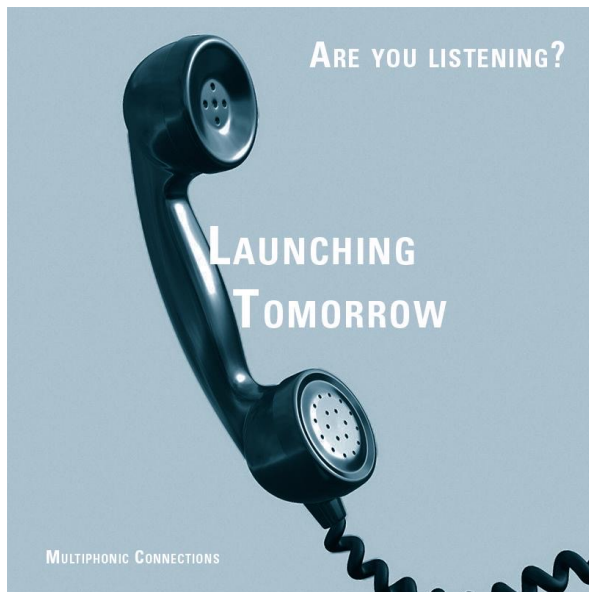


Figure 19: Multiphonic Connections social media launch poster.



Figure 20: Multiphonic Connections social media poster with telephone number.

Participant Outcomes: (via voicemail messages) wanting to hear sounds from space, such as a black hole ([Item 8](#)), the sun, moon and clouds ([Item 9](#)), sounds of collectivity beyond that which is spoken, and sounds of different time scales such as geological time ([Item 10](#)), paint drying ([Item 11](#)), sounds from the body to identify problems or unrecognisable pain ([Item 12](#)), the sound of a cybernetic mind dreaming ([Item 13](#)) and sounds of a cheese hamburger and other foods ([Item 14](#)). (Via social media): ‘This is brilliant @empathyloading 🙌 I want to hear the sounds of sleep, geological sounds (shifting of tectonic plates, volcano eruption, earthquake). My son recorded his own request as a voicemail’ and ‘Snail digesting it’s [sic] food’ (Empathy Loading, 2020 b).



Figure 21: Abbreviated voicemail message from Multiphonic Connections.



Figure 22: Abbreviated voicemail message from Multiphonic Connections.

Reflections and Analysis

The ‘Speculative Listening’ workshop was repeated and reworked into a different format because of the COVID-19 pandemic: now titled Multiphonic Connections, it became an interactive telephone experience.

The interactive telephone experience was created using an Interactive Voice Response (IVR) or ‘Auto Receptionist’ service via Voice over Internet Protocol (VoIP) from zoom.us.⁷¹ This phone system does not need a landline but uses the internet to receive calls. It enabled everything to be automated: participants entered numbers into the keypad to access different parts of the telephone experience,⁷² which was narrated by a synthesised voice using TTS with tools from Replica Studios (Replica Studios, n.d.). This time, the ‘inaudible audio’ soundtrack was divided into different categories (see below), developed partly from the conditions that limit human hearing, which we discussed at Tate Modern, and partly to add more interactivity to

⁷¹ For more information on Zoom’s VoIP Service (See: Zoom Support, 2021).

⁷² See Appendix B for full transcription of the narration of the interactive telephone experience.

the telephone experience.⁷³

On dialling the telephone number, participants were given the following introductory audio message ([Item 15](#)):

Thank you for calling Multiphonic Connections
We make the in-audible audible.

Press 0 for your initial listening calibration exercise
Press 1 for sounds from inside the human body
2 for far, far away, cosmic sounds
3 for sounds from deep below you
4 for sounds made by animals
5 for sounds that manifest very slowly
6 for sounds that happen extremely fast
7 for very quiet sounds
Press star at anytime to return to this menu
And 9 at anytime to leave a voicemail message

Participants were invited to start the telephone experience (although they didn't have to) through an 'initial listening calibration exercise' ([Item 16](#)). This functioned to focus the participants on the telephone audio rather than any other stimulus they may be encountering. The instructions for this exercise were inspired by deep listening practices by Pauline Oliveros (2005) and were developed from the previous workshop, Speculative Listening II, at which we discussed these ideas.

Due to the remote nature of this work, participants no longer physically created their ideas with materials but left voicemail messages responding to the same provocations as before ([Item 17](#)). Interestingly, participant responses exhibited the same poetic nuances as in the physical workshops. They contributed ideas that

⁷³ I would like to note here that through this practice project I learnt that music audio does not have very high quality over telephone or VoIP services since they incorporate technology that compresses sound into a limited frequency range that is appropriate for voice communication, but which limits the quality of music.

emerged from a perspective of co-creation that spanned different ecologies, time and space, human and non-human entities, with ideas of collectively and sensitivity towards the lived experience. The physical workshops and telephone experience, although in different formats with varying specific details, were united in that they were all catalysed by a soundtrack I curated to inspire participant ideas. I believe that this encouraged participants to conceptualise ideas that specifically departed from a point of sound and sounding.

Giving Voice to Synthetic Sonics

Location: Online / Zoom.us

Premise: In collaboration with Anja Borowicz Richardson, we devised this workshop as part of an open call by the Royal College of Art's Student Union for student-led workshops for other students.

Participants: thirteen MA students from the Royal College of Art. Participants were self-selecting via an email call out by the RCASU, and the workshop was advertised with a poster image (Figure 23).

Participant Outcomes: see description below.

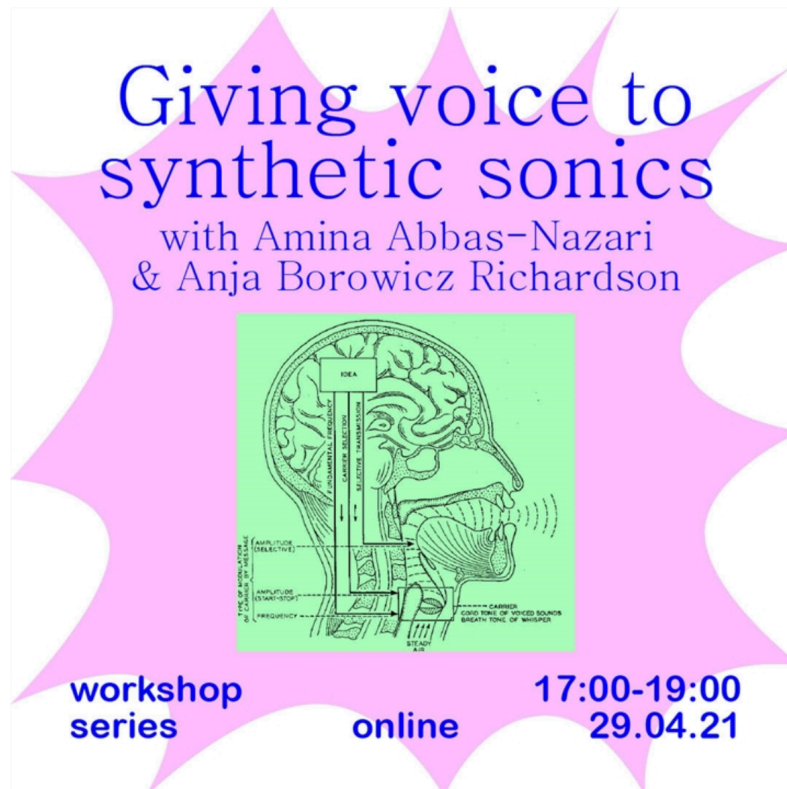


Figure 23: Poster for Giving Voice to Synthetic Sonics. Features schematic from (Dudley, 1940)

Reflections and Analysis

‘Giving Voice to Synthetic Sonics’ was an online workshop/performance, developed and facilitated in collaboration with Anja Borowicz-Richardson. It had a different structure, emphasis and concept from ‘Speculative Listening’, which I will describe. However, it serves as a helpful reference point to validate the findings of the methodological development.

We started this workshop by sharing resources and tools for voice synthesis and manipulation that are readily available and easy to use.⁷⁴ We also provided links to artworks that might inspire or be relevant to the workshop activity. All these materials were accumulated in a shared Google Drive (See: Abbas-Nazari & Borowicz-Richardson, 2021), and participants were given time to experiment and

⁷⁴ These included AI voice cloning apps, text-to-speech generators, voice synthesisers/modifiers and free audio editing software.

explore these resources as they wished. The group was led to think through ideas of plurality and polyphony of voices and of perspectives. We questioned collective ways of knowing (i.e. a chorus) versus hegemonic systems like Amazon's Echo voiced by Alexa. The workshop was framed by exploring what it might mean to protest with a synthesised voice in a sonic way. Participants used the resources and tools to create synthesised versions of their voices and modify or digitally process voice and speech sounds. The workshop culminated in the group collectively improvising a synthetic poem/performance around the theme of 'sounds of protest', as understood in the broadest sense ([Item 18](#)).⁷⁵ The aim was to test the expressive potential of synthetic voices. The workshop was open to all RCA students, and no prior knowledge of sound and synthetic media was required.

Despite the emphasis being placed on the sound of protest and the sound of voice, participants were still inclined to work with speech and words to explore their ideas. Another participant used noises that resembled sounds from weapons. One participant chose an intervention using a predefined 'Asian' synthesised voice to grapple with the current lack of accent and dialect options available to represent different people and cultures. While participants may have interpreted 'voice' in quite a broad sense, I feel they were induced to work more with words, rather than with vocal sound, because of the current limitations that synthesised voices have in terms of their sounding.⁷⁶ Additionally, I feel that 'Giving Voice to Synthetic Sonics' revealed that there is still a conceptual gap in understanding and working with the voice as sonic material, since all the participants were seen to default to using words or noises without spoken information for their contributions to the collective performance.

⁷⁵ During this performance part of the workshop, we found that it worked best to have all our cameras switched off, removing visual stimulus to foreground listening but also to lessen the potential for feeling too self-conscious to fully participate.

⁷⁶ See also Chapter 2 'Trying to Teach an AI to Sing'.

The 'Speculative Listening' workshops allowed ideas to emerge from a co-creation perspective, whereas 'Giving Voice to Synthetic Sonics' did not. Although the improvised performance at the end of the workshop produced a collective sounding of designed synthetic and synthesised voices, the individual voices themselves were neither conceived nor conceptualised as being co-created or 'intra-active' (Barad, 2007).⁷⁷ I realised that for this to be achieved, the workshops need to be initiated and rooted in sound and sounding. Here sonic thinking becomes embedded into its structure, and this practice process prevents the default to an ocularcentric way of thinking, working and making (Figure 24). For example, the 'Speculative Listening' workshops were catalysed by the provided soundtrack of 'inaudible audio' (Figure 25). 'Giving Voice to Synthetic Sonics' could have achieved similar results to 'Speculative Listening' by providing related auditory illustrations such as vocal 'choralities', described by Connor (2015) as including 'chants of protest, demand or celebration found in political and sporting circumstances', for example. This hypothesis is also supported by the 'Multiphonic Connections' practice work. In this participatory, interactive work, people were guided by the same soundtrack as in the 'Speculative Listening' workshops. However, instead of making models to represent their ideas for novel sonic experiences, they were asked to verbally describe them in voicemail messages.

⁷⁷ See next Chapter for more discussion



Figure 24: Participant blowing into plastic tube from Speculative Listening I. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys.

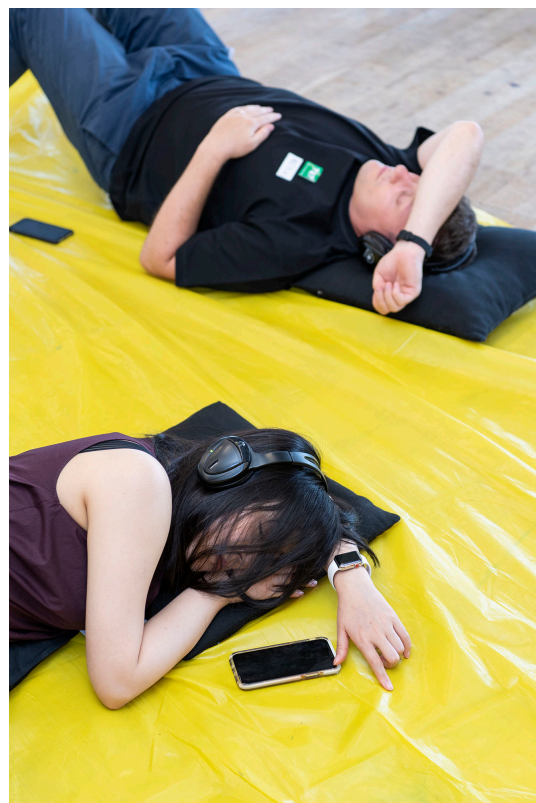


Figure 25: Participants listening to 'Inaudible Audio' soundtrack. Summer School Teachers' course in collaboration with Tate London Schools and Teachers team, 2019. Tate Modern. Photo © Tate, Joe Humphrys.

Nevertheless, 'Giving Voice to Synthetic Sonics' was beneficial in helping me consider how conversational AI systems could potentially possess a collective voice and operate more as a choir, or chorus, rather than being singular both in their representation, and how this is reflected and reinforced in their vocal sounding. These ideas of polyphony are developed further in my PhD methodology via Eidsheim's theory for investigating voices in conversational AI systems (see following chapter).

Speculative Listening IV Workshop

Location: Leeds University, Art & Design School

Premise: Invited by Louise Wilson, Lecturer in Art and Design at Leeds University

Participants: three undergraduate design students. Participants were self-selecting via an email to the department from Louise Wilson who had invited me.

Participant Outcomes: a device to mask sound when someone becomes overwhelmed/overstimulated, a device to transform tactile interactions into sound (Figure 26), communicating with animals/nature.



Figure 26: Participant making their Speculative Listening device to transform tactile interaction into sound from Speculative Listening IV. Tim C Huang.

Reflections and Analysis

This workshop was the fourth and final ‘Speculative Listening’ workshop I completed as part of this PhD research. In previous workshops, I was very excited to hear the participants’ ideas; however, on this occasion I did not experience the feelings of surprise and delight I had before. Deeper examination of this feeling revealed that this was not due to the impression left by the (brilliant) participants’

ideas, it was a result of already being able to anticipate how the workshop would unfold. In this respect, I concluded that there were no additional methodological findings from this workshop, and that future workshops could be repeated effectively and achieve similar outcomes.

Final Reflections and Conclusion

As observed and described above, participants' ideas reflected and expressed event-based proposals with complex, co-created 'intra-actions' (Barad, 2007) between time, bodies, matter and non-human animals. Using a soundtrack encouraged thinking with and through sound, to imagine novel sonic interactions. Via sonic materiality participants were prompted to contemplate ideas, issues or devices that were neither static or isolated, which relates more to objects or images. Participants felt very confident in expressing entangled ideas with very little prompting: on reflection, this is quite a difficult task.⁷⁸

In the 'Speculative Listening' workshop series, the sound and sounding of participants' designed ideas was purely imagined, highlighting how the devices' resulting sound does not need to be made audible to enable sonic-based thinking. The artefacts that participants made acted as a vehicle to encourage sonic thinking within and while designing and imagining. They were tools to allow participants to articulate how their body would be able to hear their chosen imagined sounds. Here, sonic thinking becomes a means to render intersectional new materialist-based theory tangible and experiential. This is discussed at length in the following chapter, which concerns the theoretical development of the methodology.

⁷⁸ Having conducted and facilitated many workshops, design briefs and participatory activities during my career (e.g. See: Phillips & Abbas-Nazari, 2022) I was intrigued to find that participants independently articulated and generated such imaginative and complex concepts. My usual strategies for encouraging speculative and imaginative idea generation involve the creation of a well-defined speculative scenario for participants to work within, providing a lot of additional idea generation, and/or reassurance to participants that concepts they unconfidently consider to be 'silly' or 'ridiculous', are still valid.

While these workshops have been about the materiality of sounding and listening with technology in a very broad and general sense, my focus for the forthcoming practice and research in the thesis becomes specifically about the sound and sounding of voices in conversational AI systems. Using what I have learned from these workshop experiences, I aim to maintain the poetic, ecological and social sensitivity of participant outcomes. Future practice works apply the methodology developed through the workshops, but instead of using found sounds to drive the imaginative process, I concentrate on using designed sound. I will invert this process of using found sounds to imagine speculative designs and now create vocal imaginaries illustrated through speculatively designed sound. From this point onwards, the practice projects take a more distinct shift, focusing on vococentric sounding. The methodology is explored further and tested through two case study practice projects, one for human voices, one for synthesised voices (See chapters 5 and 6).

Chapter 4: Methodology - Towards Speculative Voicing

Introduction

This chapter explores the workshop series I conducted (detailed in Chapter 3) through existing theory and practice to contextualise my sonic speculative design methodology theoretically. This methodological approach, through practice, opens new vocal possibilities by incorporating theory and perspectives from sonic thinking and speculative design to activate Question 1 of my thesis research: How can thinking with and through sound develop a sonic speculative design methodology? The first part of this chapter describes this by utilising and synthesising theory from the fields of speculative design (Dunne & Raby, 2014) and sonic thinking (Voegelin, 2014). It then moves to an intersectional stance, from which I introduce further theory from new materialist ideas by Eidsheim (2015; 2019; Schlichter & Eidsheim, 2014) to situate the sonic speculative design methodology specifically in relation to voice. Finally, the sonic concept of polyphony is introduced to formulate a methodological framework with four conditions for Speculative Voicing.

Having established that the current understanding of voice employed in the domain of AI is harmful and marginalising,⁷⁹ the development of this sonic speculative design methodology aims to critique current data-driven voice profiling in conversational AI systems. It provides an alternative framework for speculative designers working with voice, that does not rely on profiling human or imagined individuals, borrowing from lessons learnt during the series of workshops. The following chapters address Questions 2 and 3 of my thesis concerning how this methodology can be applied, where the Speculative Voicing Framework is implemented and tested through case study projects.

⁷⁹ See Introduction and Chapter 1

Introduction to Speculative Voicing

As the workshop series initiated, my methodology applies sonic thinking (Voegelin) to speculative design (Dunne & Raby) to evolve a sonic speculative design methodology, incorporating previously unexplored ideas from sound and music to speculative design.

The designers Franinovic and Serafin (2013) highlight how sound is a neglected medium in design. Meanwhile, they advocate how 'sonic interaction design' can stimulate new areas of research and practice within the design field. Traditionally, this field privileges visual media over media relating to the other senses. This follows what sound studies theorist Jonathan Sterne (2003) terms 'the audiovisual litany', indicating the way sound and sonic phenomena are overlooked, thus reinforcing an ocularcentric status quo. Equally, as Candela and de Visscher (2023) note, 'sound design' as a discipline conventionally brings to mind sound for film, advertising and acoustics in spatial design. The authors note that 'sound is suited for prompting questions, destabilising that which is thought to be stable, and for re-examining what we think we know', which is also instrumental in this PhD research. This thesis investigation is situated in the field of design and is highly concerned with sound as a primary medium, but ultimately it could be considered vococentric (Chion, 1994). I aim to speculatively design sound to create new possibilities for the sound and sounding of voices in a conversational AI context. Attending specifically to material qualities of vocal sound allows sonic-centred thinking to take precedence over visual reasoning and comprehension, which I will expand on later.

Speculative design, popularised by the designers Anthony Dunne and Fiona Raby, is employed as a dominant component of the research method and methodology for this thesis investigation, as it is already acquainted with the fields of design and technology as a tool to question itself (Dunne & Raby, 2014, p. 35). The practice aims to 'critique, and challenge the way technologies enter our lives and the

limitations they place on people through their narrow definition of what it means to be human' (p. 34) and 'offer alternatives that highlight weaknesses within existing normality' (p. 35). Through my research, I explore how the field of AI relies on profiling to understand voices in conversational AI systems and show how this is limiting. Using the methodology described, I aim to challenge current understandings of voice within AI and advocate for speculative designers to explore new possibilities when working with human and synthesised vocal material. Thus, understanding voices as material and having materiality could be a new reference point for understanding voices in conversational AI systems. I compare the disparity between current vocal profiling and an intersectional, materiality-based framework. Using the framework, I 'critique and challenge' by 'highlight[ing] weaknesses' (Dunne & Raby, 2014, p. 35-35) through revealing and resisting the way voices are currently profiled, understood and designed in conversational AI systems.⁸⁰

Practice and theory that combines sound and fiction or speculation is already a mode of working in sound, music and composition. For example, theorist Kodwo Eshun (1998) describes the potential of sound to tell stories about humans and technology. Referencing a UR record,⁸¹ Eshun narrates it as 'an object from the world it releases' (p. 07[121]). Here, Eshun describes how 'sonic fiction' can conceptualise alternatives and new possibilities, which aligns with this research project.⁸² Another example is the composer Maryanne Amacher's (1930-2009) unrealised opera work *Intelligent Life* (1980 -), which uses sound to orchestrate a speculative narrative about augmenting human listening capabilities with the use of technology (See: Cimini, 2019). Nevertheless, this approach remains under-explored in design (Oliveira,

⁸⁰ This is specifically addressed in the 'Analysis' sections of Chapters 5 and 6, which detail the two case study practice projects.

⁸¹ Underground Resistance (UR) is a collective of techno music producers and musicians from Detroit, Michigan, USA, working together since the late 1980's (See/listen: (Gaviny, 2011))

⁸² I believe the power of incorporating elements of speculation prevents works from becoming too literal, which obscures the potential to liberate our imaginations and instead results in rumination on existing problems.

2016).⁸³ However, there are similarities between sonic and speculative design practices, which I aim to synthesise. For example, the sound writer and researcher Salomé Voegelin (2014) notes the fictional qualities of sound and its potential to produce ‘sonic possible worlds’, as she calls them, in her book of the same title. She describes sound as a speculative venture: ‘It is neither a representation of an actual event nor the construction of a possible event, but is an event in all its possibilities’ (p. 32). Furthermore, an understanding of sound as ‘an alternative world, that allows us to nontrivially reconsider the status quo of what we pragmatically refer to as actually real’ (p. 32). Speculative design equally aims to ‘design for how things could be’ (Dunne & Raby, 2014, p. 12) and ‘entertain [...] possibilities for an alternative world’ (p. 92). Synthesising theory from these two fields can also address some of speculative design’s contemporary criticism.

Speculative Design meets Sonic Thinking

As speculative design⁸⁴ has grown and become recognised, it has come under scrutiny, as rightly and understandably so, for failing to imbue an ethos that aligns with contemporary critique. Oliveira and Prado (2015) state that speculative design and design fiction projects rarely incorporate the voices or have much awareness of marginalised people, while claiming to be critical of mainstream and neoliberal values. The authors note that work produced in the field could often be considered racist, classist and colonial. Projects all too often portraying ‘dystopian’ futures or alternative realities naively fail to consider that they may resemble situations that are in fact currently being experienced by people, especially those living in, or

⁸³ Sound has appeared as a medium in a few projects from the speculative design field but remains under-explored. Use of the voice as a key medium can be seen in *Across the Sonic Border (Variations on 50Hz)* (Abbas-Nazari, 2014), also Marguerite Humeau’s *Back, Here, Below, Formidable* (2011) where the artist attempts to unearth the sound of extinct animals by reconstructing their vocal organs including the lungs, trachea, larynx, vocal folds, mouth and nose (Debatty, 2011). Also, Calum Bowden’s project *Calls of Duty* (2016), *New Organs of Creation* (Burton & Nitta, 2019) and *Our Friends Electric* (Superflux, 2017).

⁸⁴ Speculative design, critical design and design fiction are terms often used interchangeably, although they differ slightly. See also: Figure 27. I use the terms design fiction and speculative design interchangeably in this chapter, as used by the original authors. In the thesis, generally, I use the term ‘speculative’ or ‘speculation’.

originating from, non-Western societies (Oliveira & Prado, 2015, p. 50). In other words, what could be considered dystopian fiction for some may be a reality for others. This critique is essential to acknowledge, as part of this PhD deals with negative impacts of vocal profiling in conversational AI systems. Nevertheless, I wish to explore these themes by utilising speculative design in my method and methodology. By confronting this quandary, through my research and practice I strive for a revised speculative design to better negotiate these issues via an intersectional position.

Furthermore, I believe speculative design has foundational principles which can allow it to evolve to provide an intersectional position to critique contemporary AI technology. In *Speculative Everything: Design, Fiction and Social Dreaming*, Dunne and Raby call for a 'shift away from the top-down mega-utopias dreamt up by an elite; today, we can strive for one million tiny utopias each dreamt up by a single person...we need more pluralism in design, not in style but of ideology and values' (2014, p.8-9).⁸⁵ However, as Oliveira and Prado (2015) show, this ambition of speculative design is failing to come to fruition. I believe a revised speculative design methodology, which takes an intersectional position, needs to be broadened to incorporate those working speculatively, but who may not label themselves as speculative designers, to move towards this original intention. Therefore, in this research I use 'speculative designers' to describe a manner of conducting creative practice which is akin to speculative design, not necessarily a particular type of practitioner or profession.

In response to the issues highlighted by Oliveira and Prado (2015), Oliveira, a founding member of the Decolonising Design platform, proposes a decolonial

⁸⁵ This quote was originally articulated in Dunne, A. (2009) One million little utopias. In: Onkar Kular (ed.). *Accept no other imitations*. London: Royal College of Art, Design Interactions.

approach that combines theories of design fiction and Afrofuturist⁸⁶ sonic fiction (Eshun, 1998), in his 2016 article 'Design at the Earview: Decolonizing Speculative Design through Sonic Fiction'.⁸⁷ Oliveira's work has been highlighted for this research because it aligns with an exploration of what sound might offer speculative design, especially when considering how the field could grow and mature critically, which I also hope to offer, and I feel is necessary. Oliveira (2016) calls for more attention to how sound is designed and how the sonic narratives of designed artefacts produce, mediate and convey listening practices (p. 44). Oliveira builds his proposition of 'decolonising at the earview' through an understanding of 'sonic fictions design futures coming from the eyes and ears of the other' to build the basis of 'theories and experiences of those alienated others' (p.51). Oliveira describes how projects should have awareness of, and offer perspectives of, people who live differing realities. Oliveira's examination focuses on listening as a method to decolonise the field of design fiction. In contrast, the methodology adopted in this thesis aims to contribute an approach oriented from the perspective of the sounding and the materiality of voicing, which I will describe.

Thinking about the materiality of sound, including voices, is to understand that sound is not discrete; it fills negative space, oscillates between bodies and interacts with all aspects of its surroundings. For example, imagine a small room with a table, and next to the table a chair – 'thinking about them in visual terms makes them separate objects, with a clear name and meaning, but what is between them and how can we rethink this world from this in-betweenness', as Voegelin (2019) says in a interview on BBC radio. Here, Voegelin expresses that thinking with and through

⁸⁶ Afrofuturism is a cultural aesthetic, philosophy of science, and philosophy of history that explores the developing intersection of African diasporic culture with technology. It was coined by Mark Dery (1993) during interviews with Samuel R. Delany, Greg Tate, and Tricia Rose.

⁸⁷ The concepts and disciplines of sonic fiction and design fiction emerged in the late 1990s / early 2000s and have their roots in the literary discipline of science fiction, but sonic fiction developed from an Afrofuturist perspective. Both fields have continued to expand; however, they have primarily remained within their own or neighbouring fields of study.

sound, or 'sonic thinking', differs from an ontology and epistemology based within a visual modality. Recognising how emphasising sound and sounding in the 'Speculative Listening' workshop series allowed individuals to easily conceptualise this co-created 'in-betweenness', I would suggest a different approach to Oliveira that asserts that there is no 'other'. As Eidsheim (2019) describes in her vibrational theory and practice approach, sound as 'a continuous vibrational field [contains] undulating energies (flesh, bones, bodies, ligaments, teeth, air, longitudinal pressure in a material medium, molecules and much more)' (p. 8). This stance does not oppose Oliveira's recognition of 'others'. Instead, it proposes that self and 'others' are enmeshed and cannot easily be separated but are in constant relational co-creation. In other words, this takes an intersectional stance that works against the duality of binaries. As will be further described, Speculative Voicing, as a methodology, takes up this intersectional position for speculative design practice via sonic thinking.⁸⁸ Speculative Voicing calls attention to the multiple realities of different people, people as multi-dimensional and different technological and ecological systems as entangled and actively contributing to each other's lived experiences. These ideas started to be exposed during the workshops, described in Chapter 3. With this further theoretical development, the practice will enact these concepts in the case study projects detailed in Chapters 5 and 6.

Developing Speculative Voicing as a Methodology

Speculative Voicing is a venture to merge the sonic with speculative design to form an intersectional sonic design methodology. Recognising that this way of working is perhaps intuitive to me because of my prior experiences,⁸⁹ the workshops I conducted enabled me to articulate, understand and teach others how Speculative Voicing as a methodology might be enacted and utilised. Working with young

⁸⁸ Martins (2014) notes a lack of speculative design practice-based works that address intersectionality, stating that most of the research in this area has resulted in 'purely textual' outcomes.

⁸⁹ See Chapter 2.

people allowed me to consider how speculative design practice, which deals with ethical implications of emerging science and technology (Dunne & Raby, 2014, p. 12), must evolve appropriately with emerging and future societal changes.

The title of the series of workshops, ‘Speculative Listening’, reflects how my research enquiry was initially interested in ‘listening’. As my research deepened, I noticed that there was a gap in the literature⁹⁰ enabling me to address my research themes from the perspective of ‘sounding’, and that this route was further supported by my prior experiences as a singer, the theory detailed in this chapter, and my findings from the workshops. Via an engagement with the materiality of sounding, participants demonstrated that this methodology enabled them to speculate and imagine new intersectional sonic experiences.

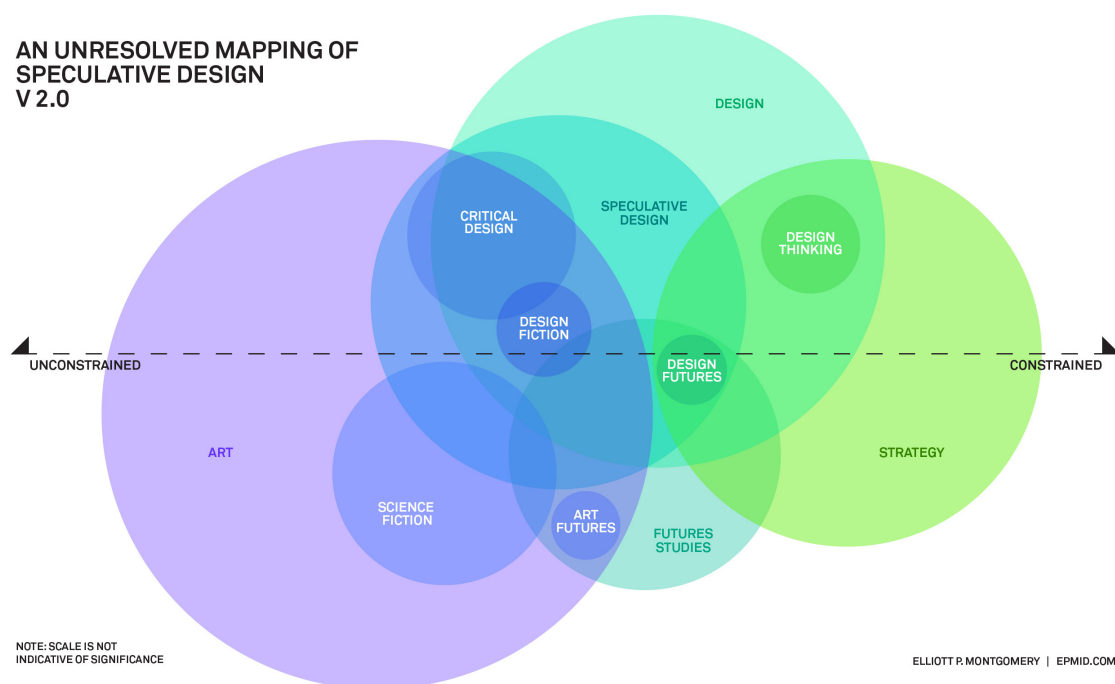


Figure 27: An Unresolved Mapping of Speculative Design V2.0. (Montgomery, n.d.)

The design researcher and strategist Elliott P. Montgomery’s (n.d) diagrammatic

⁹⁰ See Chapter 1.

exercise of An Unresolved Mapping of Speculative Design V2.0 (Figure 27) aims to situate speculative design and related fields of research and practice within a broader collection of practices and their terminology. I have tentatively placed Speculative Voicing within this modified diagram (Figure 28). By contributing sonic interaction design (Franinovic & Serafin, 2013) and sonic fiction (Eshun, 2018) to the diagram also, I situate and recognise these related or neighbouring fields of study to Speculative Voicing as a concept, and provoke further exploration from a wider range of theorists and practitioners into the fold of speculative design to develop the field. As the revised diagram suggests, sonic interaction design is located closer to the ‘constrained’ end of the axis and I would situate ‘Speculative Voicing’ as more ‘unconstrained’ and positioned close to sonic fiction.

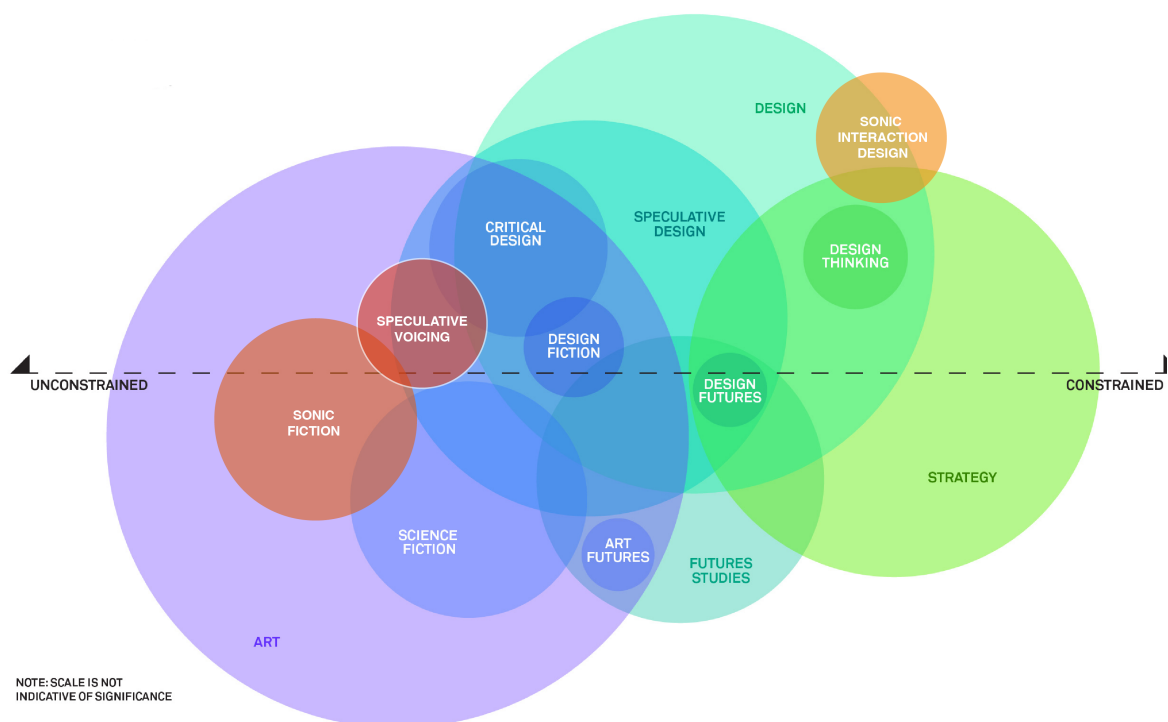


Figure 28: Speculative Voicing in An Unresolved Mapping of Speculative Design V2.0. Adapted from (Montgomery, n.d.).

The key signifying facet of Speculative Voicing is that this methodology means that sounding ‘voices’ only emerges from a perspective of intersectional co-creation, i.e. not in isolation or in a vacuum-like state. Here, ‘polyphony’, as a sonic concept,

can develop Speculative Voicing, also maintaining the curiosity and sensitivity towards the lived experience and ecological thinking, from a more-than-human and entangled perspective, that the ‘Speculative Listening’ workshops yielded. This proposition could be imagined using a choir or chorus as a metaphor within this research, which now moves to centre on voice. As a result, a voice does not constitute one person but is polyphonic – it is, and contains, many voices that all contribute to a whole.⁹¹ In contrast, from a position of self/other, this defines clear boundaries, binaries and categories, which this research and its methodology hope to negate. Polyphony, as an understanding of plurality in vocality, provides an intersectional position and also a form of resistance. As Katherine Meizel (2020) notes, in relation to researching identity through multivocality in singing:

Multivocality – as the enactment of multiple vocal ways of being – can figure as a sonic negotiation of intersectionality, a strategic intervention that supports the political use of voice against multiple, intertwined systems of oppression [...] multivocality can be a form of resistance (Meizel, 2020, p. 17).

Polyphonic Materiality in Practice

In the two forthcoming case study projects, I enact and test the affordance of the Speculative Voicing methodology by designing vocal sound and sounding in conversational AI systems. Here, vocal sound from a materiality and polyphony perspective becomes a design material, and this design is enacted through sonic thinking. This move also addresses an aim of my research, to reorientate the discussion of vocal profiling in conversational AI systems towards sounding. The methodology of Speculative Voicing aligns with Steph Ceraso’s (2022) query, provoked by their investigation of AI voice persona design company VocaliD, which:

[raised] questions about what a more equitable future for vocal technologies might look/sound like. Though I don’t have the answer, I believe that to understand the fullness of voice, we can’t look at it from a single perspective.

⁹¹ See Chapter 2, where I previously discussed and contextualised these ideas.

We need to account for the entire vocal ecology: the material (biological, technological, financial, etc.) conditions from which a voice emerges or is performed, and individual speakers' understanding of their culture, race, ethnicity, gender, class, ability, sexuality, etc. An ecological approach to voice involves collaborating with people and their vocal needs and desires – something VocaliD models already. But it also involves accounting for material realities: How might we make the barriers preventing a more diverse voice ecosystem less difficult to navigate – especially for underrepresented groups? In short, we must treat voice holistically. Voices are more than people, more than technologies, more than contexts, more than sounds. Understanding voice means acknowledging the interconnectedness of these things and how that interconnectedness enables or precludes vocal possibilities.

How might we advance an understanding and further exploration of voice that this research promotes, echoed by Ceraso? Having identified that a key concept in my PhD research is the notion of the voice as polyphonic material, Connor's (2001) brief mention of the 'phonomorphic'⁹² voice intrigued me and prompted me to think more about defining *how* a voice can morph. I wanted to explore further how vocal materiality might be shaped and sculpted and by what means this could be achieved in conversational AI systems. This was informed by my experiences as a singer and my theoretical grounding in Eidsheim's *Sensing Sound: Singing and Listening as Vibrational Practice* (2015), a contemporary reading of singing and listening with solid links to the feminist philosophical movement of new materialism. Although Eidsheim makes only light reference to new materialism in the book, it is very much present, and it is beneficial to establish this relationship more clearly for this thesis investigation in trying to merge theory and practice from the field of music with design practice.

New materialism as a philosophy posits that consciousness does not have to be a prerequisite for matter to have agency in the world (Barad, 2007). Barad's neologism

⁹² Etymologically, the term signifies voice or sound (*phone*, phono) with a form or shape (*morphe*, morphic).

‘intra-action’ challenges the notion of ‘interaction’. Interaction ‘assumes there are separate individual agencies that precede their interaction, the notion of intra-action recognizes that distinct agencies do not precede, but rather emerge through, their intra-action’ (p. 33). Notions of new materialism, or ‘vital materiality’, described as the ‘active participation of non-human forces in events’ (Bennett, 2010) fits naturally with a theoretical analysis of sound, which Eidsheim references.⁹³ These vibrant and permeable views on the nature of sound are echoed in contemporary thinking in sound studies. As Voegelin (2014) observes, ‘The sonic thing is not through its autonomy but is its action as interaction, creating not itself but the event of the moment, the aesthetic moment of the work and of the everyday as the commingling of what there is together rather than through deduction and adding up of what there is apart’ (p. 98). The ‘Speculative Listening’ workshop series made these theoretical views evident in practice.

A body of literature links sound and new materialism, especially in sound studies. However, Sterne (2012) notes that sound studies predominantly focuses on *writing* about sound. Eidsheim’s (2015) text was chosen for this thesis investigation because it focuses specifically on sound and music *practice*. Eidsheim’s position as a musicologist and classically trained singer is beneficial in describing the sound and sounding of voice within this practice-led research because it provides a theoretical contextualisation of my vocal practice. The theory gained from Eidsheim’s ideas is used to explore the conjunction of voice and its polyphonic sonic potential within conversational AI systems.

Eidsheim (2015) uses vibration, a material property of sounding and listening, to create a new materialist-based theory of singing and listening from a multi-sensory perspective. As Eidsheim describes, the text is ‘concerned with the material

⁹³ Feminist new materialist scholars, including Bennett (2010) and Barad (2007), often use concepts from music, sound and vibration to explain their theories (James, 2019).

relationship between humans and things, for which the practice of vibration is both metaphor and concrete manifestation' (p. 16). I initially considered working with vibration in my research and practice, and I explored this by creating a wearable accelerometer to pick up vibrations from my body when making vocalisations.⁹⁴ It was built from a piece of conductive fabric wrapped in felt (Figure 29) and attached to a microcontroller, similar to an Arduino, which transcribed the analogue movement from the conductive fabric into digital data (Figure 30). Plugging the microcontroller into a computer, the data could be read via the Arduino plotter function in the Arduino computer application. When testing the simple sensor device, although it was possible to generate data it was difficult to maintain accuracy because of general movement from the body interfering with the 'vibration' intended to be monitored. For example, I wanted to measure vibrations from my cheeks when sounding a sustained, held vocal tone. However, unspecific movements from my head created massive variations in the 'vibration' data. Although vibration is a constant phenomenon, the sensor I made had great difficulty picking up vibration data when placed more than a couple of centimetres away from my mouth.

⁹⁴ Thanks to RCA technician John Wild for their help in making this.

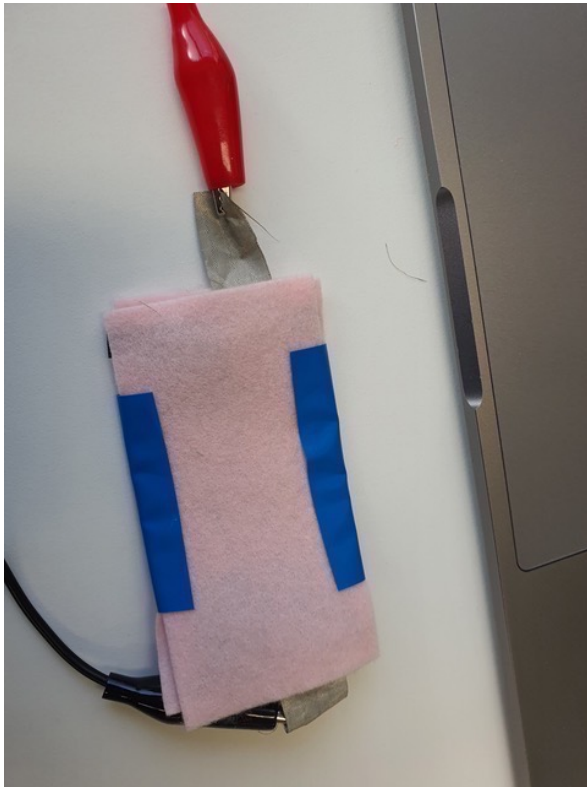


Figure 29: Vibration sensor.

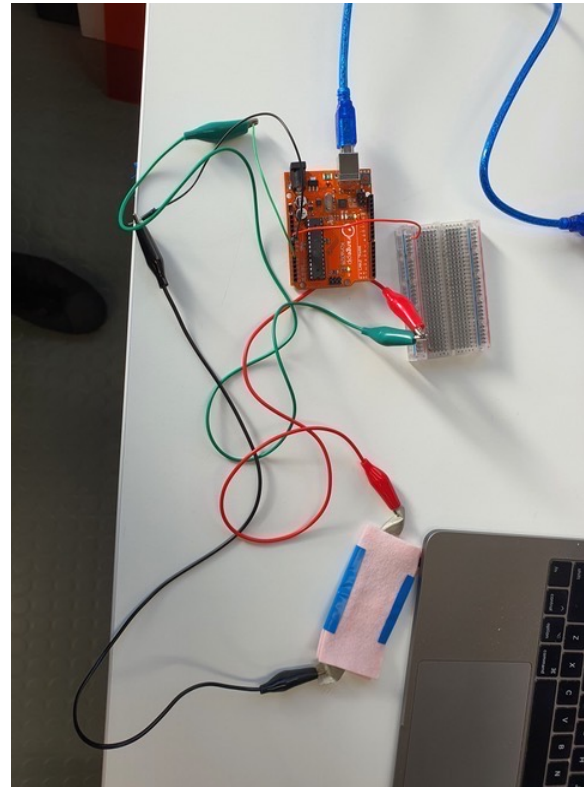


Figure 30: Vibration sensor, wired up.

I considered utilising more sensitive piezo vibration sensors.⁹⁵ Ultimately, however, I realised this venture into vibration and data collection was the antithesis of my PhD investigation, which sits more with what cannot be reduced to data. This is equally reflected in why I do not use AI, which will always aim to quantify and measure, as a principal medium in the PhD practice. Therefore, I decided to focus more on describing voice as material, its materiality and materialism (and to use this terminology). Furthermore, this position aligns with the fields of design and speculative design can accommodate it.⁹⁶ This research aims to utilise sound, as James (2019) says, to ‘be a productive model for theorizing [...] intellectual and social practices that are designed to avoid and/or oppose the systemic relations of

⁹⁵ I made (unsuccessful) contact with scientists to use Laser Doppler Vibrometry facilities to achieve something similar to ‘Non-contact Measurement of Facial Surface Vibration Patterns During Singing’ (Kitamura & Ohtani, 2015).

⁹⁶ Speculative design provides a mode of creating concepts and projects that are not material dependent or specific. Dunne (2009) says the field provides ‘[a] design approach that allows people’s imaginations to flow through objects, drawing, photographs and performances into the world around us, demonstrating at a modest scale how reality can be re-modelled, and that our own personal utopias might not be as impossible as we think’. The practice aims to render critical thought materially, using the language and structure of design to engage people (Dunne & Raby, 2014, p. 35).

domination classical liberalism and neoliberalism create’ (p. 5). In this case, I oppose these systematic relations in AI’s voice profiling practices.⁹⁷

As Kate Crawford (2021) has advocated, this research upholds the position that artificial intelligence is neither artificial nor intelligent, as it aims to be perceived. Instead, it is computationally enabled statistical modelling, materially dependent on human labour and non-renewable resources. Therefore, this research does not discuss notions of cognition and consciousness concerning artificial intelligence. However, a new materialist-based approach for this research enables a deeper investigation of the physicality and materiality of AI systems, leading to a discussion of intersectional and social themes, as the two case study practice projects explore in this thesis. In order to do this, and to advance an understanding of the voice as material and having materiality, it is necessary to formulate strategies to work with the voice in this way in speculative design practice to problematise vocal profiling in AI.

A Speculative Voicing Framework

As noted after the ‘Giving Voice to Synthetic Sonics’ workshop, participants struggled to work with voice as a sonic material and utilise its materiality. Instead, participants defaulted to working with words and the linguistic qualities of voice. This section describes the development of a Speculative Voicing Framework to guide those working to reveal and/or resist vocal profiling in, and by, AI. In order to condense the methodological theory into a practice-led tool, the concept of ‘polyphony’ is utilised to enable this.

The foundational aspects of the Speculative Voicing Framework are formed via fundamental principles of *Singing and Listening as Vibrational Practice* (2015), as

⁹⁷ Or, as Audre Lorde eloquently says, ‘The master’s tools will never dismantle the master’s house’ (1984, p. 112).

described by Eidsheim.

She says:

- (1) sound does not exist in a vacuum but is materially dependent.
- (2) the transmitting medium (for example, water versus air) and the combination of different materialities (such as the body in relation to water versus air) affect the sound's propagation and hence its actualization.
- (3) listening is materially dependent.
- 4) we can arrive at these conclusions about sounds and music only if we investigate them in a material and multisensory register (p. 49).

From this perspective, we can understand the sonics of voice as always co-created and interdependent in interaction with other matter. As Eidsheim says, this 'provides a route for thinking about fluidity and distribution that does not distinguish between or across media, and a portal for communicating beyond physical boundaries' (2015, p. 16). This statement identifies that these ideas are suitable for investigating distributed networks, such as those involved in conversational AI systems and how they are particularly appropriate to working with sonic material, including voice. Furthermore, it carves out a creative space to be imaginative and explore alternatives.

Eidsheim's observation (Point 1), 'sound does not exist in a vacuum', forms a key anchor of my argument that conversational AI systems currently do not appreciate voices as sound with sonic materiality, which contributes to harmful profiling practices. Currently these voices are distinct from the physical world from which they emanate and are embedded within. The forthcoming Speculative Voicing Framework aims to release voices from the vacuum created and maintained within conversational AI systems, to reveal and resist voice profiling. As such the voice can

no longer be insisted on as a 'fixed, extractable, and measurable "sound object[s]" located within the body', as Kang (2022) notes in discussing how the voice identification and analysis industry understands and utilises voice. These notions constrain and restrict polyphonic vocality. However, through a material understanding of voice, its materiality plays an active role in 'human ecology', concurrently 'tied to the body and entwined with the external environment, the voice exists in a complex interaction with multiple physical and sociocultural formations' (Schlichter & Eidsheim, 2014). This postulation allows this thesis to investigate sociocultural and ecological concerns via the voice. These are vital discussions to dissect, especially concerning the growing use of conversational AI systems and their profiling practices. Meanwhile, I plan to show how this opens space for vocal imaginaries, guided by Eidsheim's further points, 2-4, working with vocal sonic materiality.

The Speculative Voicing Framework provides guidance for speculative designers and practitioners working or intending to work with voice to craft creative work that builds alternative depictions of voice and voicing which do not resort to profiling practices. Instead the Speculative Voicing Framework is tasked to reveal and resist voice profiling. It is important to provide these speculative designs because through appreciating voices as sonic material and having materiality vocal profiling is rendered irrelevant and inadequate. Every sounded speculative voice provides evidence against the validity of vocal profiling, giving increased agency to marginalised voices and those who may, or have, experienced harm as a result of vocal profiling.

Working with Eidsheim's theories, in this thesis, I explore voices in conversational AI systems through four conditions of materiality which comprise the Speculative Voicing Framework (Figure 31).

As such, one voice in conversational AI systems can:

1. be embodied and co-created with other matter and/or the environment
2. be embodied and co-created with many other bodies and voices, such as in a choir
3. embody many voices, co-created with the body
4. embody many disembodied voices, co-created with the conversational AI system

These four conditions explore the polyphonic potential of human and synthesised voices and define how I work with and understand their materiality. The conditions also contrast the restrictive nature of the four auditory attributes in which voices are currently defined and designed in conversational AI systems.⁹⁸

⁹⁸ See Introduction.

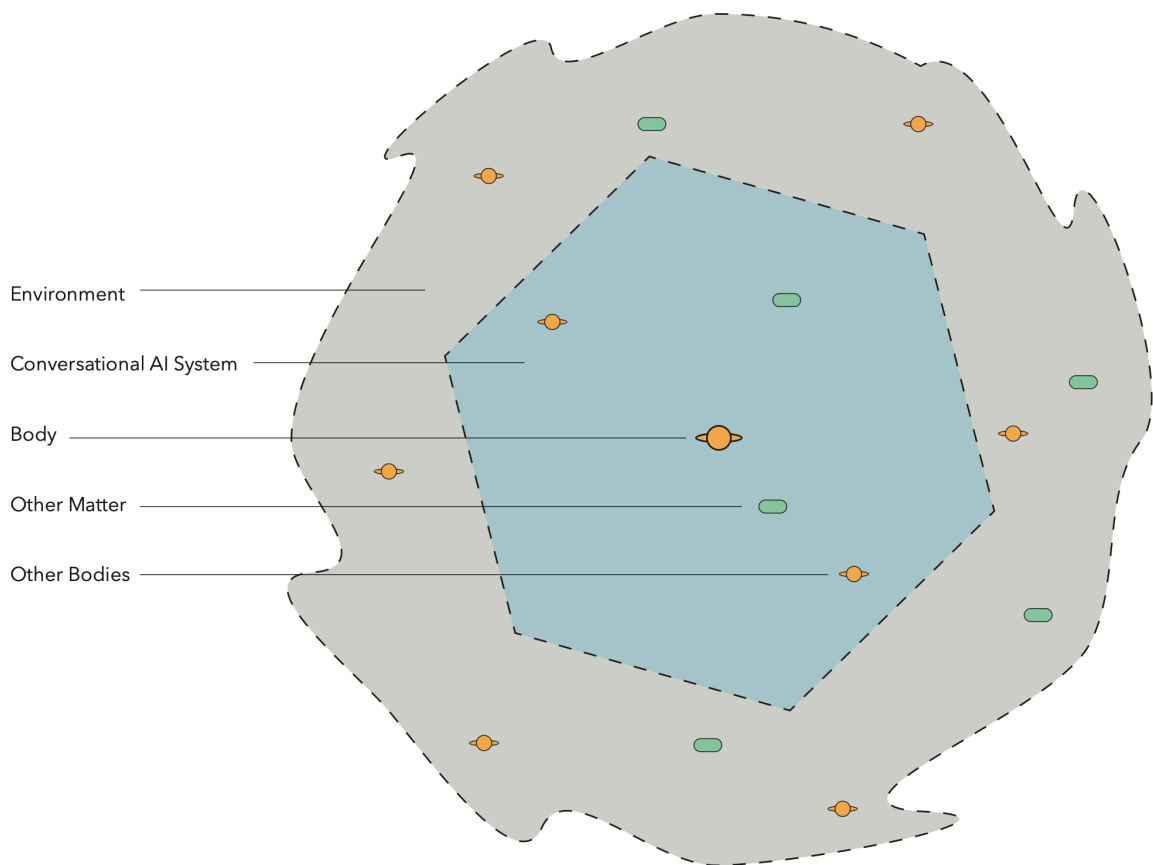


Figure 31: Speculative Voicing Framework schematic.

In my case study projects, I explore voicing as materially contingent, whereby voices are polyphonic matter. I explore how polyphony can be expanded into different constituent parts where the body, environment, matter and technology in conversational AI systems can shape the voice.⁹⁹ Speculative Voicing is employed to reveal and resist existing frameworks of understanding voices that rely on binaries, categories and taxonomies used in AI processes.

As a singer would, I work with the voice as always initiated from an embodied state. I explore the multitude of vocal abundance when co-created from this origin point and the conversational AI system. As Eidsheim (2012) says, ‘voice is always materially grounded across all points of contact, we might understand it as

⁹⁹ For example, in my practice project, ‘Acoustic Ecology of an AI System’ (Chapter 6), I use the Speculative Voicing Framework to explore how synthesised voices might morph and change their sounding to situate the voice in different physical and material environments that reveal a fuller portrayal of what it means to voice within a conversational AI systems.

corporeally enacted throughout all acts of voicing, transduction and reception' (p. 9). With the defined methodology in mind, and building upon Eidsheim's work, I traverse three layers of epistemology with which to think through and work with voice in the thesis and practice:

Voice as sonic matter – conceptualising the voice as an object of knowledge.¹⁰⁰

Voice as sonic material – an object of knowledge, that has qualities and properties which can be designed – shaped, sculpted, processed, manipulated and morphed, like any other material.

Voice as sonic materialism – an object of knowledge, that can be shaped and sculpted, in active participation with other human and non-human forces in events.

I use my experiences of experimental vocal practice to apply the methodological framework described above to directly engage with the sonic aesthetics of voice in conversational AI systems. My voice becomes a testing ground to prototype ideas and/or actively engage in the project outcomes to highlight the vocal potential of just one voice within this new framework of understanding.¹⁰¹ A lineage of female experimental vocal practitioners, as described throughout the thesis, have always pushed the possibilities of vocal expression to liberate themselves from pre-defined expectations of identity. Equally, modulation and modification of vocal aesthetics with technology, such as the vocoder,¹⁰² contributed to emancipatory Afrofuturist works of sonic fiction (Eshun, 1998). These are instances where exploring vocal

¹⁰⁰ Eidsheim says reconceptualising the voice as an 'object of knowledge' allows for analysis of the voice and voicing as a verb and not a noun (Eidsheim, 2015, p. 2-3). I question if a word other than 'object' would be more appropriate to emphasise the sonic nature of the inquiry, but I have left this as written by the author.

¹⁰¹ 'Polyphonic Embodiment(s)' (Chapter 5) directly engages with my voice in conversational AI. In 'Acoustic Ecology of an AI System' (Chapter 6), I use my prior vocal performance experiences to imagine and design synthesised vocal sound. I continue an (intersectional) feminist position. As Rosner says, women have always engaged with technology using their bodies, and she advocates for 'recognition of women's embodied practice – or bodies at all – as core contributors to engineering' (2018, pp. 435-436).

¹⁰² See Chapter 6: Acoustic Ecology of an AI System, for more discussion on the vocoder.

potential allowed for reflection on existing frameworks of oppression while opening up new creative possibilities.

Conclusion

This thesis proposes and defines Speculative Voicing as its core contribution to knowledge. Speculative Voicing comprehends the voice as sonic material and as having materiality, which can, therefore, be designed, shaped and moulded both intentionally and unintentionally. As a newly proposed methodology, Speculative Voicing builds on speculative design and addresses critique through sonic thinking and sonic materiality, taking a new materialist, intersectional stance. In the following two chapters, the Speculative Voicing Framework is applied to my two case study practice projects investigating the sound and sounding of human and synthesised voices in conversational AI systems to test the methodology, its ambitions and potential. The thesis sub-questions are addressed, asking: 2. What does applying this methodology reveal about vocal profiling in the AI era? 3. How does applying this methodology resist vocal profiling in the AI era? As developed in the ‘Speculative Listening’ workshops, I continue to produce practice using simple, readily available tools and materials, providing a bottom-up, DIY approach to making and producing work. This approach intends to sit in contrast to, and contest, AI’s top-down structuring of voice.¹⁰³

¹⁰³ For example, ‘Polyphonic Embodiment(s)’ illustrates how people voicing within conversational AI systems can have the agency to modify their vocal sound using low/no-tech devices.

Chapter 5: Polyphonic Embodiment(s)¹⁰⁴

Introduction

This chapter focuses on discussing in detail the case study practice project ‘Polyphonic Embodiments(s)’ and contextualising it within the aims of this thesis investigation. The project was created to investigate how AI transcribes understandings of voice into assumptions about identity. It features collaboration with Nestor Pestana to design and make DIY voice modification devices and AI technical development by Sitraka Rakotoniaina. [Item 1](#) contains complete details of contributors’ specific roles in this project and all the other practice works. The project speculates on how one body has polyphonic potential and how this can reveal and resist the rationale of vocal profiling frameworks currently maintained by conversational AI systems.¹⁰⁵

Project Origins

‘Polyphonic Embodiments(s)’ focuses on a particular voice profiling AI developed at Massachusetts Institute of Technology (MIT), which aims to visually illustrate an individual’s face from the sound of their voice, as detailed in the paper ‘Speech2Face: Learning the Face Behind a Voice’ (Oh et al., 2019). The authors claim that their intention is not to recreate an accurate image of a speaker’s voice but to understand how physical features correspond to sonic vocal input (p. 1). However, Figure 1 (p. 1) and Figure 3 (p. 4), included as part of the paper, seem to contradict this statement, since the documentation of the experiments shows images of whole faces, not just specific facial features. Furthermore, the results suggest a high level of accuracy when visually predicting a face. In this study, the facial characteristics correlated to vocal sounding by the AI are highly specific, including upper lip

¹⁰⁴ This practice project and some details about the work from this chapter appear in an online article I wrote for the *Sounding Out!* blog (See: Abbas-Nazari, 2023)

¹⁰⁵ The thesis also tackles vocal profiling in creating synthesised voices within AI conversational systems, detailed in the following chapter.

height, nose height and jaw width (p. 6). This study differs from other voice-to-face recognition systems that aim to identify individuals by specifying people within categories of age, nationality and gender (e.g. Nagrani, Albanie & Zisserman, 2018).

A similar paper, 'Face Reconstruction from Voice using Generative Adversarial Networks' (Wen, Raj & Singh, 2019), received notable criticism after it was presented at the Advances in Neural Information Processing Systems 32 (*NeurIPS*, 2019) conference. Alex Hanna, a trans woman and sociologist who studied AI ethics at Google (at the time of being interviewed), was contacted by *The New Yorker* after tweeting, 'Computer scientists and machine learning people, please stop this awful transphobic shit' in response to the publishing of the paper (Hutson, 2021). In the magazine article, she goes on to describe how these projects 'shouldn't exist' underpinned by four specific objections she had in particular, saying:

[...] how someone's voice resonates in the skull is not dependent on being male or female. Second, the system is likely to work better on the voices of cis people than on the voices of trans people. Third, the software's presumably higher failure rate for trans people could cause harm by misrepresenting them. Finally, the system could be used for surveillance. These objections might intersect. Hanna imagined what might happen if a trans person ended up on a most-wanted list. "I don't know if they do this anymore, but they put a composite sketch of this person on TV or social media, and then you have your old face following you around the Internet," she said – a "representational harm" (Hutson, 2021).¹⁰⁶

'Polyphonic Embodiment(s)' sought to explore voice-to-face recognition AI in conjunction with an understanding of polyphonic potential and the voice as a sonic material shaped by the body. The project invites people to consider the multi-

¹⁰⁶ The extent to which profiling people happens via conversational AI systems and how the information is then actioned is hard to discern. Often, AI and their learning systems are only exposed at the point when they are seen to fail or as a consequence of harm being inflicted, as implicated by Alex Hanna. For example, Amazon's Alexa told a ten-year-old child to touch a live plug with a penny. The AI had suggested this as a 'challenge to do', referring to the dangerous activity known as 'the penny challenge', which had been circulating on the social media platform TikTok (BBC, 2021). Here, the prompt word of 'challenge' set in motion an action with dire consequences: the AI has no conception of the harm it may inflict.

dimensional virtues of voice and vocal identity from an embodied standpoint. It calls for reflection of the relationships between voice and identity and individuals having multiple or evolving versions of selfhood. The voices' assemblage with the custom-made AI software creates a feedback loop to reflect on how people's vocal sounding is 'seen' by AI, to reveal and resist how voices are currently heard, comprehended and utilised by AI, and indeed the AI industry itself.

Human Voices and Practices of Profiling

The existing literature on voice profiling in conversational AI systems insists on the uniqueness of voices – that individuals have one voice, and a particular voice can be attributed to one person only (Singh, 2019, p. 63). This understanding renders the voice as 'an automatic and highly efficient marker of identity', as described by Nass & Brave (2005, p. 98) in *Wired for Speech*, advocating that, as the book's subtitle suggests, '*Voice Activates and Advances the Human-Computer Relationship*'. This description of the voice as unique, and its frequent comparison with the uniqueness of a fingerprint, originates in a 1911 article by the journalist R. Y. Gilbert. He postulated that 'vocal fingerprints' might be used for criminal investigation alongside actual fingerprints that are already used for these purposes. Gilbert wrote:

A criminal can shave off the ends of his fingers or burn them so that it is impossible to take a print, he may cleverly distort his features while being photographed [...] but it is almost impossible to disguise a brogue or dialect or to cover up the traces of a foreign origin in one's speech (p. 25).

Later, in 1926, *The American Journal of Criminal Law and Criminology* published an article titled 'A New Mode of Identifying Criminals' which detailed how voice analysis was able to:

distinguish between the curves produced by the voices of poets and musicians from laborers who are not appreciative of the finer arts. The method of classifying voice curves is worked out on lines similar to the

Bertillon system¹⁰⁷ of indexing fingerprints (Wigmore, p. 165).

In Wigmore's descriptions of distinguishing one person from another, we can see that voice analysis, in its earliest technological format, and profiling were coupled together. Furthermore, this was in the context of criminal punishment and upholding the law – a space where a binary of right and wrong exists, with nothing in between. It was understood as achievable because the voice was believed to be unique in its sounding. Once transcribed into an image, it formed a picture comparable to a fingerprint that could be analysed and compared to other vocal fingerprints. Voice profiling and analysis are congruent with surveillance capitalism, as documented by Zuboff (2019, pp. 245-248), but both historically, as described above, and still today find a role in criminology and law. For example, there have been instances where voice data from conversational AI systems has been used as evidence in courts of law (Whittaker, 2018). The artist Lawrence Abu Hamdan (2018) describes the use of voice analysis and profiling for border control and immigration. These themes are also explored by Pedro Oliveira in both his sound works – for example, *DESMONTE* (2021) – and writing (Vieira de Oliveira, 2021). The use of the concept of uniqueness in the context of profiling is a harmful analogy, because it insists on the traceability of a voice to a singular individual, to identify them and single them out.

This thesis argues that the concept of uniqueness should be attributed to the voices' ability to change, evolve, morph, and shape-shift. Adriana Cavarero (2005), a writer on the philosophy of voice, theorises that a human voice is 'a unique voice that signifies nothing but itself' (p. 5) and that the voice is 'the vital and unrepeatable uniqueness of every human being' (p. 7). However, Eidsheim, a professor of musicology with experience as a singer, offers a different perspective, saying: 'By

¹⁰⁷ The Bertillon system was named after the criminologist Alphonse Bertillon (1853-1914), who also invented the 'mug shot'. The Bertillon system was developed in 1883 as a filing system that contained anthropometric measurements and photography to classify and identify criminals: this was later superseded by fingerprinting (Pugliese, 2010, p. 53).

insisting on voice as event, as encultured even before birth, and as collectively projected, we can understand voice as the result of an ongoing pedagogical enterprise [...] through a series of formal and informal voice lessons' (p. 57). Here, Eidsheim observes how, rather than isolating an individual's voice as unique, humans actively mimic and learn how to reproduce vocal sound from those around them, picking up particular inflections, timbres and traits. These remarks also identify a particular problem with voice profiling, since they highlight a tension between the endeavour of a voice to be part of a collective vocal sound and sounding (e.g. accent and dialect), yet having the capability of being assigned to only one person. This conundrum could be summarised as follows: how can a voice be distinctly singular, yet collectively attributed concurrently? Or, how can voices be both convergent and divergent simultaneously? Voice profiling plays out this problem but never finds a resolution, and only feeds its impetus. It analyses vocal traits of collective groups to render a detailed picture of an individual. Nevertheless, most notably, as Eidsheim (2012) says, when consulting vocal sounding concerning identity, 'correlation is not to be confused with causality' (p. 11).

Eidsheim's distinction between 'formal and informal voice lessons' (2019, p. 57) is a valuable observation to dissect as part of this investigation because it provides an understanding of the process of how vocal sound and sounding are shaped. While Eidsheim uses formal lessons to signify voice training as part of singing, music and performance education, I would like to expand this to include how understanding the voice as material that can be shaped and sculpted, could also be considered a form of vocal lesson. Here, this research links to the context of speculative design practice: the case study projects, such as 'Polyphonic Embodiment(s)', become a tool to communicate the autonomy that voices and vocal sounding can possess. I argue, then, supported by Eidsheim's theory, that people have, and/or possess the potential to have, many voices. This is explored throughout this thesis via the defined methodology and exemplified in the case study practice projects.

Notes on the Dataset

In this section I will dissect some of the inner workings of the AI created for 'Polyphonic Embodiment(s)', namely the dataset and the decisions made during the process of making it.

'Polyphonic Embodiment(s)' aimed to recreate a version of MIT's voice-to-face AI recognition algorithm. The detailed analysis by Speech2Face (Oh et al., 2019) is, in part, achievable because MIT's system uses an unsupervised artificial intelligence. An unsupervised AI is left uninstructed to find patterns and correlations in the data provided for analysis (Goodfellow et al., 2016, p. 105). It uses a dataset that a human has not pre-labelled into categories – for example, age, nationality, and gender. It can look beyond categories and taxonomies known and understood by humans. At some level, this mitigates bias that may be imposed by human labelling of the data at this granular level. However, the collection, accumulation and maintenance of an unlabelled dataset is known to be a biased process in itself – bias is already knowingly or unknowingly present. In other words, categorising and labelling data is problematic before an AI is even brought into the equation, because using pre-existing datasets come with pre-existing biases. Birhane, Prabhu and Kahembwe (2021) investigated instances of misogyny, pornography and negative stereotypes present in commonly used, openly available datasets. They note that research by Peng, Mathur and Narayanan (2021) found that three major large-scale image datasets remain widely available through file-sharing websites, despite retractions of harmful material and content. More worryingly, long after their retractions, the datasets were used hundreds of times in published papers and continue to be used by the machine learning community in peer-reviewed research (p. 14). Voice profiling is harmful because of the reliance on data containing racist and normative assumptions, as described.

While trying to recreate the Speech2Face AI, the dilemma of assembling our dataset emerged. What might constitute an unbiased dataset? How could a representative dataset be created? More questions and more dilemmas quickly eliminated certain strategies. What about creating a dataset that incorporates as much diversity as possible? However, what does diversity, as a concept, actually mean? Alternatively, could we create a dataset that was equal – equally representing all races and genders? But how many different races or genders of people even exist in the world? Or a dataset that was representative, for example, representative of the population of the UK? However, how would the bias contained within the data available on these subjects be mitigated? Facebook recently classified faces into six different categories: pale white, white, light brown, brown, dark brown, and very dark/black, aiming to create a ‘fairer’ dataset (Hazirbas et al., 2021). How would voices and accents be categorised? How many accents even exist? The problem with even posing these questions or trying to make a ‘fair’ dataset is described by artist Trevor Paglen and demonstrated in his artwork *From ‘Apple’ to ‘Anomaly’*, exhibited at the Barbican Centre (2019 b). He says,

Every time you create a taxonomy, there's always a politics to that – because when you're creating a taxonomy, you're saying this is a range of categories that are intelligible, and it's always going to be a limited range. In doing so you're always creating a negative space, the things that are outside of that, the things that are not intelligible (Paglen & Downey, 2020).

‘Polyphonic Embodiment(s)’ seeks to challenge the notion, as Paglen describes, that people can be categorised in this manner, and in the case of this research, that this is achievable through voice recognition and profiling. Pauline Oliveros highlights the fluidity of voice and its evasiveness in relation to attempts to categorise it into binary divisions in her piece *Sex Change* (1985). In Oliveros’ ‘text score’, the written instructions position the reader as both a performer and audience, for the piece to ignite the reader to contemplate possibilities of vocal sounding and

listening. Oliveros¹⁰⁸ beautifully illustrates the multiplicity and polyphony of voice in its meditation, expression and sounding. In *Sex Change* (1985), she instructs the reader to:

Listen inwardly to the sound of your voice.

Listening inwardly to the sound of your voice changed to the opposite sex.

Listen inwardly to the sound of both voices together.

Listen inwardly as if there were many of you.

Listen inwardly freely as your voices change randomly.

Express your voices aloud.

For 'Polyphonic Embodiment(s)', it was decided not to create an original dataset but instead work with the dataset used by MIT's Speech2Face AI, known as AVSpeech (Ephrat et al., 2018). In part, this was because there was no way to answer the questions that arose. But more importantly, because this project was not about creating an intervention at dataset level, but more about the AI itself, as a complete system that tries to construct an individual's facial appearance by using data from the sounding of their voice. Therefore, for 'Polyphonic Embodiment(s)' it was more appropriate for us to use the tools described by Oh et al. (2019).¹⁰⁹

Recreating the AI

For this project, the expertise of creative technologist Sitraka Rakotoniaina was employed to (attempt) to recreate the AI, as described in the MIT paper.¹¹⁰ The AI

¹⁰⁸ Although Pauline Oliveros is primarily known as a composer and for 'Deep Listening' (2005), this 'text score' demonstrates experimental vocality in its imagined sounding. Oliveros has also written compositions for performers, including voice.

¹⁰⁹ Existing projects seek to investigate the datasets of AIs. For example, Caroline Sindors takes a 'positive discrimination' approach with the Feminist Dataset project (2017-), aiming to fashion a dataset with references purely from or about self-identifying women. A different avenue is seen in Anna Ridler's works, where she creates her dataset from scratch but avoids any issues that may emerge about racial bias since her subject matter and dataset do not incorporate humans. For example, her hand labelled dataset in the work *Myriad (Tulips)* (2018).

¹¹⁰ I initially contacted the authors of Speech2Face (Oh et al., 2019) based at MIT's CSAIL department to request to use the AI they had created, as detailed in the paper, but received no response.

created was trained using voices and their correlating faces from the freely available AVSpeech dataset, the same dataset used by MIT's Speech2Face (Oh et al., 2019).¹¹¹ The AVSpeech dataset comprises audio-visual clips 3-10 seconds long from 290k YouTube videos in which the audible sound belongs to a single, speaking person (Ephrat et al., 2018). With limited resources and computing power, 2500 samples were randomly selected from the AVSpeech dataset. Therefore, the AI created is not a replica of Speech2Face: we aimed to imitate its recognition capabilities, but in a limited sense. The samples were trained by a generative adversarial network (GAN)¹¹² image translation algorithm, where audio data is transformed into a spectrogram image. The AI then inspects the matching face and spectrogram images to find patterns in the visual data. Rakotoniaina programmed the GAN to run over 1000 epochs: during each epoch iteration the AI adjusts the weighting of discriminators and optimisers compared to a ground truth example provided. The epochs were exported every 10 seconds, creating models available for use in the project. Rakotoniaina coded an online Google Colab application called [wav2face](#) to access cloud-based computing power provided by Google's remote servers, to work easily with the newly created voice-to-face recognition AI. [Item 19](#) shows video documentation of how the Google Colab wav2face can be installed and used.

A similar project by Murad Khan and Martin Disley (working together under the name Unit Test) also sought to recreate this specific recognition AI, Speech2Face by MIT, as they described in their lecture 'Speculative Voices and Machine Learning', delivered at Unsound Festival in Krakow, Poland (Disley & Khan, 2021). Also, later,

¹¹¹ Google originally created the dataset to solve the so-called 'cocktail party problem' – enabling an AI to detect a single speaker from an audio source with multiple people speaking and background noise (Ephrat et al., 2018). This is an auditory processing ability that humans are very capable of attuning to but, until recently, was limited for machine listening abilities of user voice interfaces, such as Amazon's Echo.

¹¹² Generative Adversarial Networks were designed by Goodfellow et al. in 2014. They comprise a framework that trains two neural networks simultaneously, which compete with each other. The network dynamically learns from its mistakes and gains when generating new data with the same statistical modelling as the original statistics provided during training. GANs are a form of deep neural network and are often used for AIs that are unsupervised (Goodfellow et al., 2014).

in their work *Not I* (Unit Test, 2023). Disley and Khan were motivated to reproduce Speech2Face to understand the technological logic behind the AI's models. They describe taking an investigative engineering approach to the work to uncover the underlying assumptions and problematics of machine learning, to find ways to manipulate the AI, showing their weaknesses. In *Not I* Disley and Khan perform an 'adversarial attack' on the AI via spectrogram images.¹¹³ The duo's work draws attention to the important issue of data extraction and extrapolation for profiling that this particular AI affords. Our approaches and understanding differ significantly, largely because of the site of our intended interventions into the system. While Disley and Khan focus on manipulating spectrograms, 'Polyphonic Embodiment(s)' deals with the voice itself and its sounding before it is transferred to spectrogram image. In turn, this impacts the extent to which we each interrogate the role and understanding of the voice in AI-enabled recognition and profiling. Ultimately, our understanding of the voice more generally is affected due to our differing backgrounds – Disley and Khan are creative coders and I am a singer and designer.

Human Voices in a Vacuum

While recreating Speech2Face, the lack of actual sonic material in the AI's learning process became very apparent: the way an AI 'listens' uses image recognition. The AI's learning process in both Speech2Face and our recreation of the AI recognition algorithm correlates pixels of RGB arrays on paired images of faces and voice audio spectrograms to find patterns. Although these AIs deal with sonic data, as part of the learning process the audio recording is transcribed into two-dimensional images to create an image-based AI recognition process; these images are then further abstracted into lines of code. In this case, the AI bridges and metamorphoses sonic data into visual information and then numerical data to forge its understanding of how voices relate to faces. At no point does the AI understand

¹¹³ An adversarial attack in machine learning is a digital attack that aims to mislead the AI model with deceptive data. It is purposely created and contains hidden data to cause an AI to make an error in its prediction, which resembles a valid input to a human (Boesch, 2021).

what pitch or colours or frequency are, for example. Vocal sound is stripped of its sonic materiality and isolated into sonic data. However, through a perspective that understands the materiality of the medium of sound, as this investigation does, the voices escape the captivity of being rendered as an image, such as a spectrogram. Consequently, this is why this research argues that voices in conversational AI systems exist as though they are in a vacuum. A vacuum, an airless void with no other particles or matter contained within it, prevents air and, therefore, sound from existing. Voices in a AI vacuum render it still, trapped and suffocated.

Human Voice as Sonic Material

‘Polyphonic Embodiment(s)’ uses designed materials to shape the body and modify voice to explore polyphonic potential. However, related examples of voice modification exist in musical and cultural contexts. The most apparent and common reference for this is code-switching. In linguistics, code-switching is when an individual speaker switches between two or more languages or language varieties (Woolard, 200, pp. 73-74). Although, as Woolard says, code-switching is particular to the same ‘speech-event or exchange’, we can also think about how people shift speaking styles (including how you sound your voice) between different groups, or in different settings. For example, how people speak to their family could differ from how they speak with co-workers.

There are many instances of humans exploring the sonic material of voice. TransVoiceLessons is just one example of channels on YouTube that offer resources for trans-identifying people wishing to modify their vocal sounding to align with their gender identity. One such video is titled ‘Voice Feminization for ABSOLUTE BEGINNERS | How to Get Started Now’ (TransVoiceLessons, 2021). In the video, viewers are encouraged to mimic the presenter through exercises exploring higher pitches and vocal tones to speak comfortably with a ‘cleaner, lighter, and higher sound’. Project Spectra is described in the paper ‘Online Community-based Design

of Free and Open Source Software for Transgender Voice Training' (Ahmed, Kok & Howard, 2020). It is a particularly notable project because the creators advocate for users of their open-source app to be able to define their individual vocal sound and sounding rather than being contingent on normative gender expectations. The authors note that trans scholars describe how people must align with certain social norms to present as a particular gender (p. 6): for example, the expectation for trans-women to sound more 'feminine', with a softer-sounding, higher-pitched voice. The authors were motivated to create the app because they found that existing voice training apps tend to be developed by cisgender, white European American women and consequently that they often perpetuate normative, racialised and classed gender categories (p. 7). This precedent breaks the illusion that appearance and voice must share a strong correlation.

For trans-identifying people who undertake this type of vocal training, the body (the vocal organs and the whole bodily structure) is not a defining factor for sounding their voice, as authors like Singh (2019) would proclaim, which forms a cornerstone of vocal profiling. Trans-identifying people transcend the social-cultural pedagogical training of voice, as described by Eidsheim (2019). They produce their voice through their own motivations and autonomy to define a vocal sound suited to their gender, identity and being. Project Spectra aligns with the ambitions of this PhD investigation to actively interrogate the normative expectations embedded into vocal sound and sounding and aspires to disengage these relations to give people and their voices more autonomy. Through vocal training, or, at the very least, hearing a voice impersonator at work, an understanding can be gained that the voice can be morphed within the materiality of an individual's bodily architecture or materially co-created with other environmental matter. Here the body is an apparatus to explore the polyphonic sonic potential of the sounding of voice, amplifying the agency of voices in conversational AI systems in order to actively challenge voice profiling frameworks.

In the field of music, composer and artist Meara O'Reilly explores the multiplicity or polyphony that a single voice can spawn. In her compositions, she employs the use of hocketing¹¹⁴ to create 'pseudo-polyphony', or the auditory perception that there are more voices or sound sources than are actually being voiced. She is inspired by 'auditory illusions found in indigenous folk practices, popular music, and scientific research' (O'Reilly, n.d). As she describes, in the recording of *Musique du Burundi*, (Ocora Records, 1968):

In this traditional mourning song from Burundi, a woman uses her own body resonances to completely alter the timbre of her voice beyond recognition. Her lips function like a reed on a woodwind instrument – they are set in motion by the volume of air contained in the cavity formed by her two hands clasped against her mouth. The resultant sounds vary in pitch, timbre and volume according to the position of her hands and the tension of her lips (O'Reilly, n.d).

In this instance, the woman uses her whole body as an instrument, incorporating non-speech-related body parts to modify her voice. For computer scientists such as Singh, this polyphony or the polyphonic potential of voice, presents a problem when trying to develop 'technology for the automated discovery, measurement, representation and learning of the information encoded in voice signal for optimal voice intelligence' (Singh, 2012), as Singh notes in describing her research interest. Polyphony, as a concept, is an unexpected phenomenon for one body or one voice to present in the realm of computer science without it being considered a form of masking or disguise (Singh, 2019, pp. 15-17), to obscure or obstruct technology designed to measure, quantify or correlate. As O'Reilly describes, indigenous and folk voice practices explore polyphony, voiced by one body, for musical expression. Potter (2020) describes the importance and cultural significance of hand-made and designed 'voice-disguisers' and 'acoustic masks' in West Africa, used for storytelling,

¹¹⁴ Hocketing is a compositional technique where a melody is split into brief phrases divided across multiple voices or instruments (Moreland, 2019). The quick succession of the musical phrases between parts creates a dispersed but unified melodic line.

ritual and making audible the voices of ghosts. She says voice disguisers function ‘not simply to physically distort vocal sounds but to also manifest deities and the spirits of ancestors’ (p. 310). As Potter indicates, vocal polyphony in non-Western settings is used to explore concepts of being and identity, both personally and communally.

The text above gives a sample of normative expectations currently ingrained within vocal profiling through explorations of instances where voice is, or has become, material. They include the correlation of voice to visual appearance and gender and the neglect of non-white, marginalised voices. However, this could also be extended to other people who do not fall within distinct categories and normative expectations, such as those who have speech impairments, are D/deaf, disabled, queer, trans, non-binary or mixed-race, for example. This is important to recognise from the intersectional position this research takes. As can be seen from the above examples, these vocal practices are initiated or exist in spaces that do not, or cannot, align with current profiling practices in AI systems, which are grounded within a straight white ontology, as described by Birhane (2021).¹¹⁵ Although Birhane describes this by referencing artificially intelligent systems generally the exemplars described above show that this pattern extends to the sound and sounding of voices.

Making and Using the Devices with our AI

The devices I created and designed with Pestana utilise simple, easily resourced materials to display DIY possibilities for voice modification, highlighting the materiality of the physicality of voice originating from an embodied state. By proposing a decentralised form of manufacture in the speculative project, the intention is to reclaim autonomy for voices in conversational AI. This contrasts with and disputes the other main force in the project – the artificially intelligent voice-to-face recognition agent – a top-down system that currently dominates the defining,

¹¹⁵ For a more extensive discussion of Birhane’s writing, see Chapter 1.

describing and detailing of people by the sound of their voice.

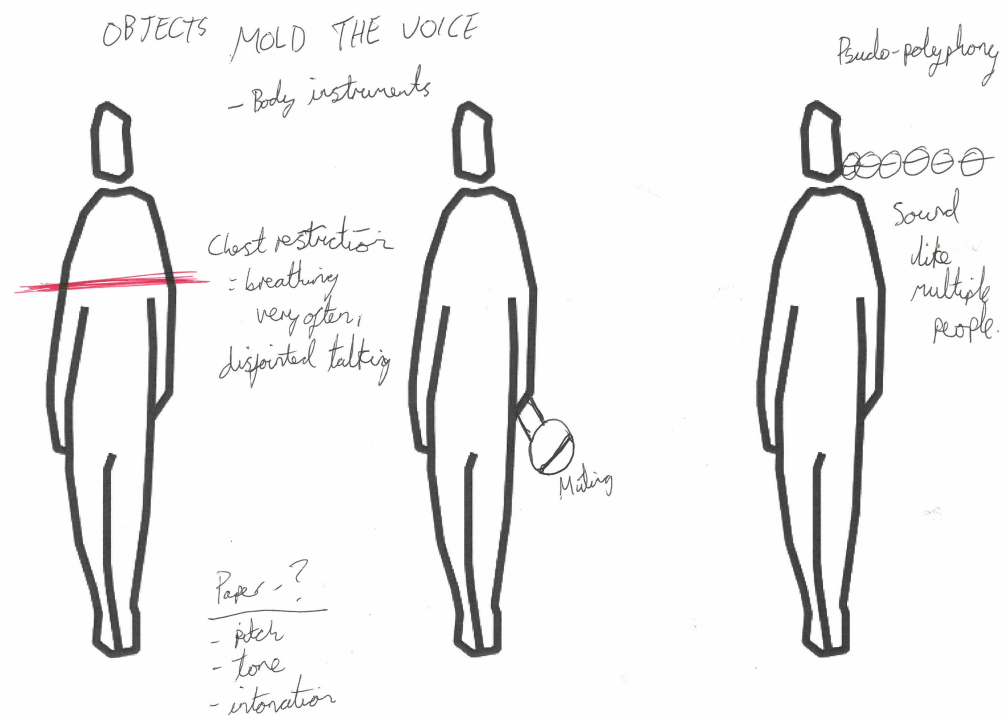


Figure 32: Initial sketch ideas for Polyphonic Embodiment(s). Amina Abbas-Nazari & Nestor Pestana.

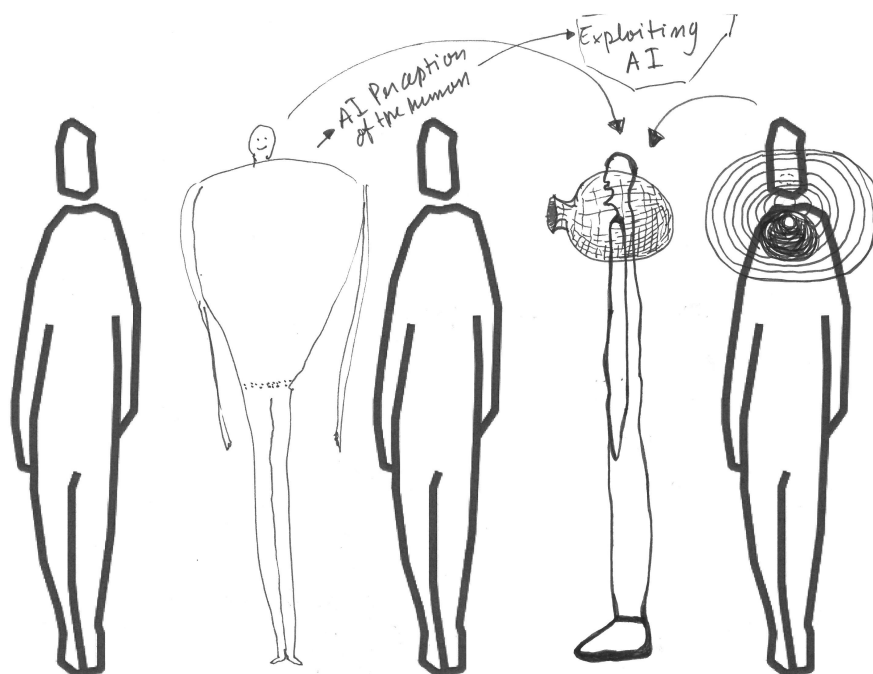


Figure 33: Initial sketch ideas for Polyphonic Embodiment(s). Amina Abbas-Nazari & Nestor Pestana.

To initiate the project, we sketched potential interventions we could make to the human body to shape bodily architecture and/or voice to modify the resulting vocal sound (Figures 32 & 33). In the next stage of development, readily available materials and objects¹¹⁶ were resourced to prototype ideas for the devices. I tested them with my vocality to see how their materiality could affect vocal sounding ([Item 20](#)). In one such example, I placed aluminium foil in contact with my lips. The interference between the vibration of my lips, the movement of air from my mouth and the thin sheet of metal changed the timbre of the sounding of my voice.



Figure 34: Testing wav2face Google Colab.

In the first instance, the wav2face Google Colab application was tested with audio from the initial material experiments ([Item 20](#)) to see the faces that were being produced (Figure 34). After reviewing the facial images and associated spectrograms

¹¹⁶ This included plastic gloves and cups, a metal bowl, a water bottle lid and exercise resistance bands. These materials were sourced from around the home and recycling bins. This was also the case with the ‘Speculative Listening’ workshops (Chapter 3) and the ‘Speculative Voicing Workshop’ (Chapter 8).

of these experiments, I realised it was necessary to minimise the sonic and acoustic variables when using the devices. This was to produce AI-derived faces most influenced by the effect the device was having and not, for example, by the volume or length of the audio clip. As shown in Figure 35, the length of the audio clips was compressed into images where the size and resolution of the spectrograms stayed the same.¹¹⁷ To enable continuity, I ensured the sound clip levels were consistent, and recorded in the same location on the same day.¹¹⁸ Also, as seen in the documentation video, I repeat the same line of text while wearing each device.¹¹⁹ The phrase, “they had no fixed values to be altered by adjectives and adverbs. He was pressing beyond the limits of his...”, lasting approximately six seconds, was obtained from *CMU ARCTIC*. This database consists of around 1150 utterances compiled by Carnegie Mellon University specifically to aid the production of speech synthesis (Carnegie Mellon University, n.d).

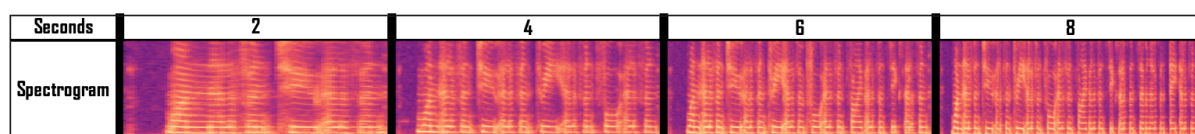


Figure 35: Timed spectrogram test for Polyphonic Embodiment(s).

After the initial material experiments, the next step was to develop these materials into voice modification devices. Returning to Singh (2019), I was intrigued by the author’s accounts of relationships between bodily characteristics and voice quality descriptors. For example, ‘the vocal tract plays a significant role in imparting the quality of ‘yawniness’ to voice. ‘Yawny’ speech is produced in the configuration of the vocal tract that widens the oral cavity and increases the tract length’ (p. 250). We initiated our voice modification devices by making an artefact to create a ‘yawny voice’ effect, resulting in Device #4 (Figure 36). Ten subsequent devices were made,

¹¹⁷ For this test, I recorded myself saying, “1 elephant, 2 elephant”, “1 elephant, 2 elephant, 3 elephant, 4 elephant”, etc, and then reviewed the spectrogram images.

¹¹⁸ Thanks to RCA Sound Studio technician Joe Hirst for assisting in helping to set up the studio and recording equipment in line with our project requirements.

¹¹⁹ Although I instigated strategies to ensure consistency during audio recording, the wav2face application is not particularly accurate or stable when producing faces of the speakers.

to which I ascribed vocal quality descriptors as noted by Singh (2019, pp. 242-251), and the materials used to achieve this effect (Figure 37).



Figure 36: Device #4 from Polyphonic Embodiment(s). Amina Abbas-Nazari & Nestor Pestana.

The assigning of voice quality descriptors was done retrospectively after making the devices, allowing the project to be led by practice and through making. In addition, while Singh uses vocal quality classing to relate to physiognomy, I was more inclined to use the terminology as subjective references to explore a single voice's polyphonic potential. As with all the practice work in this thesis, the ideas presented are not concrete examples; they are open-ended illustrations of vocal imaginaries to provoke new understandings of voice and to prompt others to explore (their) voice in similar ways. The six-second audio clips of each voice variation created by each device ([Item 21](#)), were inputted into the Google Colab wav2face application to generate facial images. The video documentation and final output of 'Polyphonic Embodiment(s)' ([Item 22](#)) shows the 10 devices we made being used and the resulting face produced by the AI we created on the left-hand side of the frame. Some images resemble my face? (e.g. Device #8). Some might be deemed more masculine? (e.g. Device #10). Moreover, some are just disturbing (e.g.

Device #4).

Order / Ref	Device Number	Vocal Qualities ?	Effect / <i>Materials Used</i>
1 / A	7	Constricted. Laboured. Weak.	Loss of airflow from voicing into an <i>extra large balloon</i>
2 / B	3	Muffled. Tinny. Buzzy.	Face and mouth smothered by <i>regular balloon</i>
3 / C	5	Stiff. Narrow. Focussed.	Cheeks squeezed in, compressing mouth by <i>dowel in a foamboard frame</i>
4 / D	9	Bright. High. Clear.	Cheeks pushed up by <i>foamboard in a cardboard neck brace</i>
5 / E	2	Biting. Forced. Breathy.	Biting down onto two <i>wooden clothes pegs in foamboard brace</i>
6 / F	8	Twangy. Metallic. Ringing.	Speaking into a <i>plastic tube connected to an expanded polystyrene ball</i>
7 / G	6	Tinny. Crackly. Resonant.	Lips and mouth pressed upon a <i>sheet of foil, held in place by a plastic pot</i>
8 / H	4	Yawny. Open. Tight.	Mouth held open at sides by <i>two metal clips in a cardboard frame</i>
9 / I	10	Nasal. Low. Flat	Nostrils held closed by a <i>metal clip in a cardboard frame</i>
10 / J	1	Thin. Coarse. Covered.	Mouth and lips pressed agitated <i>plastic straws held by a cardboard face mask</i>

Figure 37: Polyphonic Embodiment(s) table of vocal qualities, effect and materials used for device.

In the later stages of this practice project, finalised in July 2022, I discovered a similar project by Eidsheim (2012). *The Voice Box* project (1999-2012) is documented in the paper 'Voice as Action: Towards a Model for Analysing the Dynamic

Construction of Racialised Voice’, in which Eidsheim presents three different designed wearable devices to modify properties of vocal sound. These investigations specifically challenge vocal timbre as a marker of race, whereas my investigation incorporates the broader scope of profiling and the context of AI systems. In this work, Eidsheim, too, recognises the voice as material, as this research does, to question the voice as ‘commonly believed to be an unmanipulable attribute’. It is curious and reassuring that Eidsheim and I, both singers, reach similar conclusions in understanding and questioning dominant understandings of voice as potential markers for categorising identity. I hope our projects add weight and validity to recognising the voice as material with agency in its material interactions with AI.

Speculatively Voicing Human Voices in Conversational AI Systems

As previously described in this thesis, the four conditions of materiality collected under the Speculative Voicing Framework are as follows:

One voice in conversational AI systems can:

1. be embodied and co-created with other matter and/or the environment
2. be embodied and co-created with many other bodies and voices, such as in a choir
3. embody many voices, co-created with the body
4. embody many disembodied voices, co-created with the conversational AI system

I employ these four conditions to assess the potential of my case studies project, ‘Polyphonic Embodiment(s)’ to produce vocal imaginaries. In the following section, I describe why human voices in conversational AI systems should be comprehended in this manner, moving towards an intersectional appreciation of the voice with an ecological and social sensitivity. These sections of text respond to the answers to Q2 & Q3 of my research: How can applying this (Speculative Voicing) methodology reveal and resist vocal profiling in the AI era?

'Polyphonic Embodiment(s)' takes the 'Black Technical Object'¹²⁰ as described by Amaro (2019), as a starting point, since the objective of the work is not to find a more corroborative way for people to fit within normative standards imposed by recognition systems but to 'catalyse future affirmative iterations of self'. As Amaro illustrates, recognition systems can be used as mirror to observe the lens through which people are currently seen. Subsequently, fractures and inconsistencies are illuminated, making space to exist and grow beyond classifiers that contain and constrict being. Amaro's (2019) ideas are put into practice to explore vocal profiling. This thesis, and 'Polyphonic Embodiment(s)', intend to find new understandings of being that place more emphasis on the individual as part of a Whole, as ecology entangled, and not restricted to being singular and individual.

One voice in conversational AI systems can:

1. be embodied and co-created with other matter and /or the environment

In this practice project and the thesis, I highlight the voice and the way it originates from an embodied state. The sound and sounding of human voices emerge from the body, regardless of cultural, social or pedagogical learnings of voice. I refer to the sound and sounding of voices since the initial human vocal sound, created from an embodied place within the body, differs from vocal sounding. The resulting vocal sounding is a co-created formation combining factors of the body, environment and the AI system in the context of this research. Dolar alludes to this multiplicity (2006, p. 73). However, Blesser & Salter (2009) describe it more succinctly as 'proto...' and 'meta...' (p. 136). For example, a violin is a resonant enclosure that produces its particular timbre of sound, which can be described as 'protoviolin'. The violin, as a primary resonant enclosure, coupled with a secondary resonant enclosure, such as a concert hall, creates a different sounding as a result, as the 'metaviolin'. Therefore, another way to describe the sound and sounding of

¹²⁰ Described in further detail in the literature review, Chapter 1.

voices in conversational AI systems would be ‘protovoice’ and ‘metavoice’. The voice as a sonic phenomenon has materiality. Therefore, the voice in this respect is not constrained, contained or fixed but inhabits multiple states simultaneously, extending from an embodied state. With this understanding, potential emerges to shape and morph the voice through each of these simultaneous states within the conversational AI system. The potential for new sonic possibilities within these simultaneous states is highlighted to give the voice amplified agency through polyphony, from its initial conception in the body to its outer manifestations, to contest the current practice of vocal profiling. ‘Polyphonic Embodiment(s)’ explores the materiality of one voice through its exaggerated co-creation with other matter through using simple materials.

One voice in conversational AI systems can:

2. be embodied and co-created with many other bodies and voices, such as in a choir

‘Polyphonic Embodiment(s)’ is also motivated by a trend for increased vocal homogenisation as a result of factors such as monoculturalism, accent neutralisation (Aneesh, 2015) and language loss. A curious relationship also exists between vocal homogenisation and biodiversity, known as ‘biocultural’ diversity. While biodiversity is vital for animal and plant life to flourish, research links biodiversity decline to diminishing cultural diversity. Terralingua¹²¹ conducted research showing that ‘the trend in the loss of global linguistic diversity revealed by the Index of Linguistic Diversity (ILD) closely mirrors the trend in the loss of global biodiversity for the same period of time, as measured by the World Wildlife Fund's Living Planet Index’ (Terralingua, n.d.). The research points towards a broader interconnectedness between human and non-human life, for which the sound and sounding of human

¹²¹ Terralingua, founded in 1996 by linguist Luisa Maffi, ‘supports the integrated protection, maintenance and restoration of the biocultural diversity of life – the world's invaluable heritage of biological, cultural, and linguistic diversity – through an innovative program of research, education, policy-relevant work, and on-the-ground action’.

voices is an indicator of decline. However, for this investigation, that advocates for the voice as material, this also ignites a curiosity to speculate on how the sounding of voices could promote 'biocultural' diversity. It might be far-fetched to suggest the conscious and/or speculative sounding of voices to achieve this. However, via this research, I want to draw attention to the voice as having a more expansive social and ecological relationship, by virtue of its materiality, than is currently accepted in vocal profiling practices in conversational AI systems. 'Polyphonic Embodiment(s)' aims to challenge vocal homogenisation by exploring the potential vocal range one body can possess.

One voice in conversational AI systems can:

3. embody many voices, co-created with the body

Ethnologue, a catalogue of all known languages worldwide, reports that 367 languages have died out since 1950 (Romaine, 2017), and technology plays a role in this. In the context of voices in conversational AI systems, a startling exemplar is the tech start-up company Sanas who use AI technology to provide real-time voice alteration for call centre workers to make their voices sound more Western (Chan, 2022). As a result, it is anticipated that people marginalised from AI systems might be inclined to alter their voices to sound more Western¹²² to be able to more easily access, fit into, and use these types of technologies. Indeed, people with regional or ethnic American accents have recounted how they find themselves distorting their mouths to imitate Midwestern American accents, hoping to be better heard and understood by their Google smart speakers (Rangarajan, 2021). Johann Diedrick's artwork *Dark Matters* (2021), presented by Squeaky Wheel Film and Media Art Center, Buffalo, New York, also draws attention to this issue. *Dark Matters* attends to the absence of Black speech in datasets that train the voice interfaces of artificially intelligent consumer devices such as Alexa and Siri. Through an interactive online

¹²² See also: *Sorry to Bother You* (2018), the dark comedy film in which Cassius, a Black man working as a telemarketer, is advised by an older colleague to "use your white voice" to be more successful at his job (Riley, 2018).

piece and installation, the work ‘challenges our communities to grapple with racism and inequity through speech and the spoken word, and how AI systems underserve Black communities’ (Squeaky Wheel, 2021). Ironically, since speech and voice are classed, recognised, and computed as separate entities by conversational AI, as previously discussed, marginalised people experience poor usability in interaction with these systems (Koenecke et al., 2020). Yet they disproportionately experience negative impacts from being profiled by them.¹²³ However, as the examples in this chapter show, the embodied human voice is more malleable than is often understood, especially when vocal sonic identity and visual identities do not seamlessly align for an individual. ‘Polyphonic Embodiment(s)’ aims to take a deeper look at how AI ‘sees’ voices in conversational AI systems.

One voice in conversational AI systems can:

4. embody many disembodied voices, co-created with the conversational AI system

It is crucial to address how voices are understood, especially concerning notions of being and identity, as AI-mediated communication increases, as described.¹²⁴ Humans involved in conversational AI systems have no control or exposure to how their voices are understood, or how their voice data is being used. Conversational AI systems are always listening (Lau, Zimmerman & Schaub, 2018), enabled by seven directional microphones concealed by a sleek facade (Crawford & Joler, 2018). However, while they actively listen, they only passively respond. Fundamentally, humans ventriloquise conversational AI systems. Their synthesised voices speak when spoken to and need human input to learn and be trained. By interacting with a conversational AI system, a disembodied human voice becomes entangled and

¹²³ See Chapter 1

¹²⁴ Networked technology is also increasingly used to mediate human-to-human communication, such as the digital communication platform Zoom. Although not directly enabled by AI, often these types of communication platforms incorporate AI deep learning technology for real-time noise suppression to reduce background noise such as fans, dogs barking and traffic noises (Intel, n.d.). While further discussion is outside this project’s scope, I would like to draw attention to the potential relevance of this thesis research to this study area.

actively engaged in a multiplicity of roles, enmeshed into their unseen networks, enabled by comprehensive machine listening processes. As Crawford & Joler (2018) describe, a human speaker or user of a conversational AI system is ‘simultaneously a consumer, a resource, a worker, and a product’. ‘Polyphonic Embodiment(s)’ reveals how human voices might be rendered by AI while resisting the ability to be profiled.

Analysis of Polyphonic Embodiment(s)

My autoethnographic, thick-description analysis¹²⁵ of this case study practice project aims to highlight the divergence between the current understanding of human voices in conversational AI systems in comparison to this research’s exploration and potential sounding of voices, defined by the methodology and method used. The contrast between the two modes of understanding the sounding of voices builds critique by revealing and resisting vocal profiling to address questions 2 and 3 asked by my research. In order to do this, I will use Device #1 (See: Figure 38 / [Item 23](#)) as an exemplar and then discuss the experience of performing in this case study as a whole.

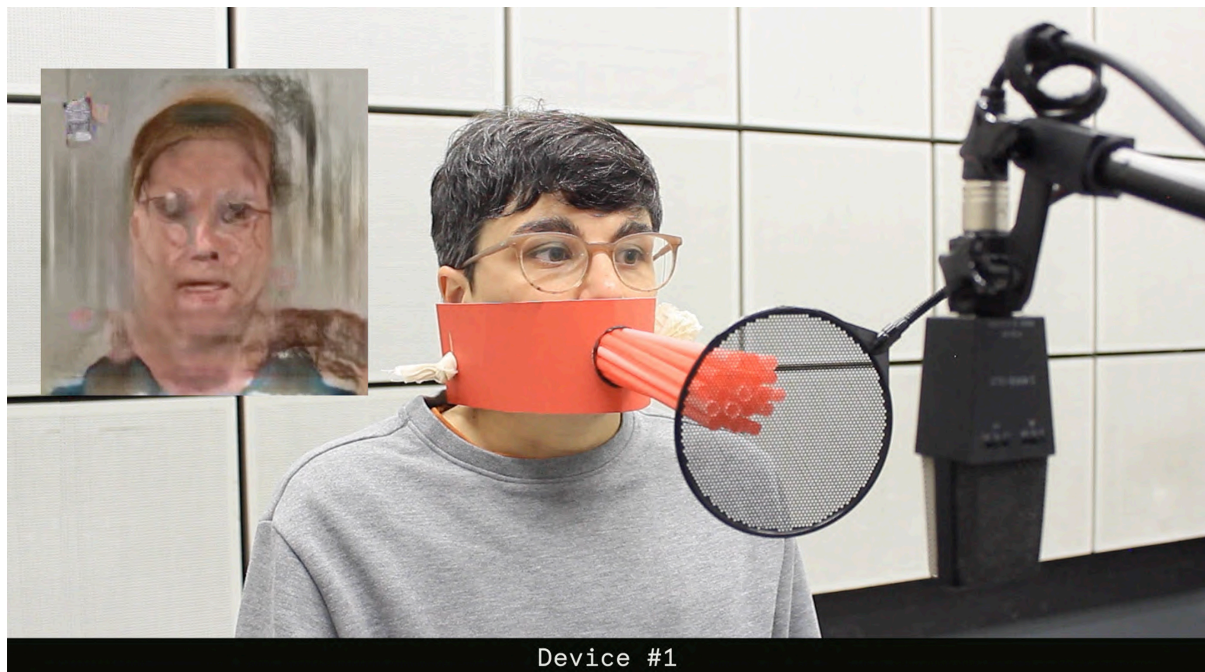


Figure 38: Device #1 from Polyphonic Embodiment(s). Amina Abbas-Nazari & Nestor Pestana.

¹²⁵ See Introduction for more information on methods.

My lips meet the plastic straws softly pressed against my skin. I begin to sound my voice, and the breathy oralisations produce high-pitched, airy whistles in between and around the smooth, shiny tubes. In this moment, my voice and cheap polymer meet, creating a new voice. Still my voice, but different. Not the voice of the plastic straws, not in duet with them either. The vibratory interference makes an unfamiliar voice, only enabled in our co-creation. I endeavour to collaborate and work in conversation with the material, now enveloped with my body, making slight adjustments to the tone, volume, and timbre of voice to agitate the previously inanimate material. Only in this cooperative voicing between my body and the body of straws can this novel voice emerge. This voice is a speculative voice.

Perhaps, like Nick Cave's *Soundsuits*, I adorn and ornament the voice to render it beyond the identity I visually and sonically represent. I listen for a reciprocating response from the plastic tubes, finding a vocal resonance that can animate them to the state of becoming where we work in chorus. This vibratory interference is not reserved for exaggerated instances, as performed here. It demonstrates a magnified version of every sound and sounding of voices as they emerge in interaction with other matter and materials. When we use the voice, especially to speak, it is easy to think that it is created and made by the voice box or larynx, isolated within this small region of the body. In 'Polyphonic Embodiment(s)', the body as a whole is recognised as an architectural form that shapes the breath to modify voice in co-creation with the encircling environmental matter. Through this morphing of my bodily and facial architecture, with the use of the devices, I am made more aware and attuned to the role of my flesh, muscles and bones in creating the voice that is mine. Moreover, I now comprehend my voice as material, not bound by the body in its sounding, in the creation of voices which are also equally mine.

In some of the first tests with the AI for 'Polyphonic Embodiment(s)', the

experience was like ‘seeing faces in clouds’. I was convinced I could discern features about my face in the images and that elements of the images resembled my physical appearance. This was short lived though, especially when I was suddenly presented with an illustration of a horribly disfigured face (Figure 39). There was a desire to understand the AI and comprehend as the AI comprehended. What could the AI perceive which was beyond my perception? This elusiveness encapsulates an AI, yet it is alluring to enquire about. As is the case with many AI models created through unsupervised learning, the outcomes are difficult to decipher and require novel methods to enable them to be ‘explainable’ (e.g. Crabbé & van der Schaar, 2022). With the introduction of every new voice modification device we had made, morphing my voice, I was presented with another AI-generated face. Was this face another version of myself, looking back at me? This face was a contortion, assembled from all the faces intricately known by the AI but unknown to myself, humanly incapable of holding the catalogue of images in my mind.



Figure 39: Example wav2face AI generated face.

The AI's resulting faces were ingested into a one-dimensional assemblage of pixels. As a human, I could trace the outline of an unsettled face, yet clearly the AI only saw RGB dots, unable to define a foregrounded face from a trivial background setting. In turn, this drew attention to the adjustments researchers at MIT must have made to the images and how much influence the AI was given in defining a face shape as part of its predictive outcome. This technique also creates a vacuum, whereby the materiality of voice has been constricted to a facial outline and this imposed, artificial boundary separates it from the rest of the material world. Our less sophisticated version of MIT's AI did not crop faces against a consistent white background as part of the pre-training procedure.¹²⁶ I found the unexpected assimilation of my newly found face submerged into a shimmering spectral effect strangely reassuring. One of the ambitions of this project was to dissolve the perimeters between binary categories and taxonomies, self and other, voice and environment. Coincidentally, this simplified AI, which did not crop the faces, helped to visualise just that. Facial features of gender and age became fluid and ambiguous. Without a discernible inside versus outside, or self in opposition to other, perhaps here lies an opportunity to re-connect ourselves to a broader ecological entanglement, where the individual exists but in relation to and co-creating with the Whole. 'Polyphonic Embodiment(s)' tries to make the hard edges and bounding boxes of categorisation, labelling and naming, often reinforced by AI-based reasoning, more permeable and porous.

Conclusion

Projects such as Mozilla's Common Voice aim to address usability issues of conversational AI systems by 'teach[ing] machines how real people speak' by building an open-source, multi-language dataset of voices (Mozilla, n.d.). However, Sterne and Sawhney (2022) describe how this approach, to mitigate problems of bias

¹²⁶ The cropping of faces against plain backgrounds has its origin in the development of film and photographs through the use of 'Shirley Cards', which today plays a role in racial bias in AI facial recognition systems (See: Camp, 2020).

or ethics, ultimately generates more data to feed the ‘will to datafy’, which supports the ‘wills to classify and identify’. In other words, profiling is merely being repeated and reinforced, with increasing attention on those disproportionately negatively affected by AI recognition systems. Therefore, it is imperative to reveal and resist current vocal profiling systems so that the fundamental understandings of vocal sound and sounding can be broadened and appreciated.

In ‘Polyphonic Embodiment(s)’, with the use of simple everyday materials to meet high technology, I found a queer poetic anarchism in creating many identities constructed in the ‘eyes’ of this AI – blurring the boundaries of self and other, resisting AI’s desire for datafication of bio and bodily markers. This project splits the voice and its correlating identity that was previously conceived as singular into many. Body and identity become multiple, obtaining polyphonic potential to be a chorus. One voice is many. ‘Polyphonic Embodiment(s)’ as a practice project allowed for an exploration of freedom of (vocal) expression of relational sonic/visual identity. This highlights how voice cannot be constrained by profiling but can, in fact, be used to observe of profiling frameworks at play, while grappling with notions of being and identity.

As with both case study practice projects, I do not proclaim that the voice *should* be designed, but that it can be considered as a material which *can* be designed, from a speculative position, to explore broader social and intersectional issues. The simply designed voice modification devices work with bodily architecture and exaggerate its materiality, considering it as a malleable instrument. In turn, this reveals the normative assumptions held within AI’s construal of voice and its relationships to facial image and identity analysis.

Having argued the polyphonic nature of an embodied human voice, with this understanding I aim to break the reinforcing cycle of voice profiling between the

understanding of human voices and the design of synthesised voices.¹²⁷ In the following chapter, I will apply the Speculative Voicing Framework to synthesised voices to find alternative ways to sound them for revealing and resisting vocal profiling by AI.

¹²⁷ See Introduction and Chapter 1 for more information on this relationship.

Chapter 6: Acoustic Ecology of an AI System

Introduction

This chapter examines voice profiling of synthesised voices in conversational AI systems and how it relies on normative expectations. The common practice of the female gendering of synthesised voices, combined with their use of language, portrays women as subservient, inferior, and in positions of servitude (West, Kraut & Chew, 2019). More vocally diverse voices are neglected or not offered at all as synthesised voice (Baird et al., 2017). The case study project '[Acoustic Ecology of an AI System](#)' (Abbas-Nazari, 2020) is an interactive online audio experience providing a vocal imaginary to reveal and resist AI vocal profiling. In the project, I apply the four framework conditions of the Speculative Voicing methodology. Employing the framework supports new ways to conceptualise and sound synthesised voices that are more aligned with the material world they interact and operate within.

In this chapter, I first describe some previous approaches to this research project, which have now been set aside in favour of the Speculative Voicing methodology. Then I discuss synthesised voices and current profiling practices to design the sonic aesthetics of these voices. I explore how the design and aspirations of synthesised voices exist as though they are in a vacuum and how this further contributes to negative profiling. I discuss synthesised voices, their use as sonic material, their materiality and the application of the Speculative Voicing Framework. The technical considerations of the project are detailed, followed by analysis and discussion of the project findings and implications.

Notes on the Project Title / Iterations and Development

The title of this chapter and the associated project refer to previous approaches to this thesis investigation, which were disregarded due to developments in the research project. As the title suggests, the research was previously framed around

and utilised research from acoustic ecology (Schafer, 1977).¹²⁸ Alongside this, theory from media archeology, in particular, Jussi Parikka's (2015) writing, was used to contextualise the project.¹²⁹ The final project output, [Acoustic Ecology of an AI System](#) (2020), was published on the online platform Attune, created by the Research and Waves collective when I was asked to respond to their provocation, 'Can words be neutral?' (Attune, n.d.).¹³⁰ A short essay accompanying the online work explored themes within the project and was further contextualised around poetry, particularly writing by Ihde (2007) and Berardi (2018). Acoustic ecology, media archeology and poetry, as the basis of a methodology and conceptual framework, were disregarded because of the need for more theoretical support and understanding of voice, which ultimately is the nucleus of this thesis research. Residual embers of embodiment and phenomenology (Ihde), sound and ecology (Schafer), technology and materiality (Parikka) and aesthetic exploration of voiced communication, i.e. poetry (Berardi), are still present within this research. However, *Sensing Sound: Singing and Listening as Vibrational Practice* (Eidsheim, 2015) provided a way to condense these multiple strands of exploration under one theoretical framework and methodology that foregrounds a contemporary and highly relevant understanding of the sound and sounding of voices for this context.¹³¹

Text-to-Speech, Synthesised Voices and Practices of Profiling

Contemporary synthesised voices are designed by creating persona profiles based on imaginary human people aligning with and appealing to a company's customers and consumer markets. *Wired for Speech: How Voice Activates and Advances*

¹²⁸ Acoustic ecology, an idea originated by R. Murray Schafer, suggests that we try to hear the acoustic environment as a musical composition and, furthermore, that we take responsibility for its composition (1977, p. 205). The practice's study relies heavily on field recording and soundscapes as composition.

¹²⁹ Media archeology is a theoretical enquiry into the material and materiality of media cultures from a historical perspective.

¹³⁰ This case study project was developed over a series of outputs: public presentations at two symposiums: the *(Un)Sound Barrier Symposium*, Royal College of Art (19th June 2019) and the *SPARC Symposium: Land Music*, Music Department at City, University of London (12th-14th September 2019). These dialogic occasions helped me refine the presentation of the work.

¹³¹ See Chapter 4 for further information on methodology.

the Human-Computer Relationship (Nass & Brave, 2005) contains many examples of experiments involving people interacting with synthesised spoken dialogue systems, such as conversational AI systems.¹³² Authors describe how the main facets considered in constructing a synthesised voice persona typically include gender, personality, accent, ethnicity and emotion. Nass and Brave (2005) note that people attribute human characteristics to synthesised voices, primarily because humans are the only living beings to communicate with speech. They say that by providing technology with characteristics associated with being human, such as human-sounding voices, people apply social rules that are similar to those expected in human-to-human interactions. The authors advocate the benefits of building on intrinsic human-to-human interaction to design synthesised voices to be perceived to be human, or human-like. With no visual accompaniment, assumptions about synthesised voices and ‘whom’ they originate from are largely based on pitch, pitch range, volume and speech rate of the sounded voices (Nass & Brave, 2005, pp. 34-36).¹³³ For example, an adult woman’s vocal pitch typically ranges from 165 to 255 Hz (Watson, 2019). Most outputs of voice assistants are created synthetically, even when modelled on a usually female human voice, and most people perceive female-sounding voices as cooperative (West, Kraut & Chew, 2019, p. 96). An Amazon representative interviewed by *Business Insider* said that the company’s research found women’s voices to be more ‘caring’, which, in commercial terms, means that devices with female voices are more likely to be used for assistance and purchases (Moynihan, 2020).

The use of conversational AI systems has grown considerably, so the demand for synthesised voices has increased. Between 2008 and 2018, the frequency of voice-based internet search queries increased 35-fold (West, Kraut & Chew, 2019, p. 92) and

¹³² Also known as and described in the book more broadly as ‘voice user interfaces’ (VUI).

¹³³ In contemporary synthesised voices, these parameters can be adjusted using Speech Synthesis Markup Language (SSML) in TTS applications (Alexa Developer, n.d.).

has continued to rise in recent years. Companies now offer to construct bespoke synthesised voices. Replica Studios is relatively forward thinking in its approach to the sound design of synthesised voices, offering varying 'styles' of voice such as 'light-hearted', 'polite' and 'serious'. However, this is because their intended markets also include the gaming industry, where more deviation from normative expectations is afforded. Replica Studios enables this by working with voice actors who record many hours of speech. The actors read from a corpus of text, which an AI learns to mimic, including 'speech patterns, pronunciation, and emotional range' (Replica Studios, n.d.). The corpus may include frequently required phrases and voice responses; however, specially created speech synthesis databases exist that provide 'phonetically balanced' phrases that are already labelled allowing easy extraction of the spoken sonic data (Carnegie Mellon University, n.d.). The sonic data is segmented into separate sounded units of speech, such as phonemes, of which there are ~44 unique sounds in English. This tonal code can be assembled into sequences that form words and sentences, which forms the process of text-to-speech synthesis.¹³⁴

Synthesised Voices in a Vacuum

The aesthetics of synthesised speech in conversational AI systems have always remained conservative, as they aim to imitate human voice communication. Authors from Google Deepmind (Oord et al., 2016), writing about recent developments in speech synthesis for conversational AI systems, maintain that synthesised speech should sound as natural as possible, and that a synthesised voice is intended to be

¹³⁴ There are two main types of text-to-speech (TTS) speech synthesis – concatenative and statistical parametric. Concatenative speech synthesis uses a database of speech waveforms annotated with prosodic and phonetic contextual information assembled to create words and sentences (Hunt & Black, 1996). Statistical parametric speech synthesis involves automatic machine selection of appropriate units by averaging sets of similarly sounding speech segments (Black, Zen & Tokuda, 2007). This is made possible by The Hidden Markov Method (HMM). The TTS, or STT system, registers a phoneme (the smallest element of speech), and there's a certain probability of which phoneme will follow. HMM uses probabilities to determine the arrangement of phonemes to form words and their most likely order. Most voice recognition systems today use HMM to understand speech (Brown, 2021). The database of vocal sounds for TTS can be created by purely electronic means or by pre-recorded human speakers.

indistinguishable from a human voice to a human listener. Authors describe how this process, currently achieved through TTS, intends to mimic how humans produce speech in their 'speech production-related organs' by computational means. However, by positing that synthesised speech is a computational re-creation of the human anatomy that produces speech sounds, this assumes that the voice is situated purely within a body. Voices in conversational AI systems currently exist within a metaphorical vacuum, in which the actual sound and sounding of voices as materially dependent is ignored. Meanwhile, this research argues that the voice is equally embodied in co-creation with a wider ecology and material world.

Synthesised voices are created to be conceptualised as anthropomorphic, with human-sounding discourse, and are perceived as a personification of AI (Abercrombie et al., 2021). As Krejci (2018) points out, they are given recognisably human names like 'Alexa', which are female-like in gender; however, where she is from, and what her interests, beliefs, or ideologies are, is hidden. While synthesised voices may simulate a face in the human imagination, they emanate from devices which are minimalist in their design and provide no visual clues to locate the particular vocal sound source.¹³⁵ Despite emanating from tangible electronic devices, synthesised voices in these conditions are acousmatic,¹³⁶ since their sound design and production offer no sonic cues to establish a physical origin. The current sound design of synthesised voices is sonically flat and acoustically unassuming, and they provide no auditory cues to situate them in real or imagined environments. They are presented as though seemingly untethered to the physical and material world in terms of space and time, which human voices act within. This notion also aligns with

¹³⁵ Bruder (2020) calls for user experience designers to acknowledge the actuality of their design. He believes designers should not obscure the reliance on human labour and non-renewable resources with minimalist, sleek, shiny surfaces when designing devices such as Amazon Echo.

¹³⁶ The term acousmatic originates with Pierre Schaeffer's concept of 'musique acousmatique', deriving from Greek legend that describes Pythagoras's disciples listening to him while behind a curtain (Schaeffer, 2017). Pierre Schaeffer, writing in 1966, was an electronic music composer and the term was coined at a time when recording technology was first emerging. This development made it possible to sever the link between sound and its source.

my argument that these voices exist as though they are in a vacuum.¹³⁷ Space, time and environment are facets that define acoustics and describe environmental features that a voice can illuminate – including the spatiality, shape, volume and materials that constitute the setting. These factors also shape the sounding of a voice through its materiality. For example, heavily furnished libraries suppress the chatter of people concealed in their corners, and audio recordings of choirs in cavernous concert halls sound as though they are situated in these spaces. However, these sonic qualities are not auditorily present in the synthesised voices of conversational AI systems. In contrast, the context of this materialist-positioned research suggests that a voice is continuously shaped through material factors, including technological systems. A co-creation between embodied voice, environment and disembodied voice produces a multiplicity of voices in one instant and yet also emanates as a whole. Synthesised voice design does not currently account for this co-creation afforded by the materiality of voices as a sonic material.

‘Acoustic Ecology of an AI System’ is an exploration of revealing and resisting vocal profiling, as posited by my research questions. Applying the methodology of Speculative Voicing to synthesised voices in conversational AI systems, using acoustic and sound design, adds a sense of the material world that fundamentally underpins these digital systems. This perspective aims to push back against the vacuum-like current conditions that reinforce vocal profiling. Advances in synthesised voices in conversational AI systems have concentrated on intelligibility and naturalness (Sutton et al., 2019). However, this project investigates how sound could be used to reattach and locate disembodied synthesised voices in space, time, environment and architecture to produce vocal imaginaries with alternative narratives of the technology with which we now, using our voices, so intimately

¹³⁷ Instead, perhaps they are designed this way to intensify the qualities of an acousmetre, identified by Chion (1999, p. 24), as being ‘all-seeing’, holding the ‘ability to be everywhere, to see all, to know all and have complete power. In other words: ubiquity, panopticism, omniscience, and omnipotence’.

interact.¹³⁸

‘Acoustic Ecology of an AI System’ intends to address how conversational AI systems present themselves versus what they conceal. It identifies how the sound design of synthesised voices also shrouds the wider material topography it is co-created with. The explorations provide a means to navigate and conceptualise conversational AI systems while critiquing their vocal profiling practices. My intention is for this project to recognise the normative assumptions being embedded into conversational AI systems while encouraging greater speculative exploration of these types of voice-enabled technologies.

Synthesised Voice as Sonic Material

In 1939, the Voder (Voice Operation DEMonstratoR) was unveiled by Bell Labs at the New York World’s Fair: it was the first electronic synthesis of human speech (Dudley, 1940). The electronic piano-like device, with a human operator, composed strings of segmented sounds into speech, punctuated by an electronic hiss to mimic human breath. The device still forms the basis of speech synthesis today. It marked the birth of the mechanisation of voice. In turn, it unravelled speech as tonal code that could be assembled into sense and substance by humans. A video uploaded to YouTube (Roemmele, 2016) presents a recording from an original demonstration of the Voder during a live radio broadcast. The presenter describes the device using electrical filters, attenuators and frequency changers to produce 20 basic sounds. Intelligible speech could be synthesised from various combinations of these sounds, controlled by a skilled female operator manipulating a keyboard and foot pedal. Using the phrase ‘she saw me’ as an exemplar, the tonal emphasis is shifted between each word of the phrase to answer the following questions: ‘who saw you?’ – ‘SHE saw me’, ‘whom did she see?’ – ‘she saw ME’ and ‘well, did she see you or hear

¹³⁸ More thoroughly explained in the ‘Making of...’ section below.

you?’ – ‘she SAW me’.¹³⁹ Here, the Voder and its operator expose the sonic system of voice at work in language-making. This demonstration shows speech reproduced synthetically, via coded forms, and exposes it as a material which can be sculpted and shaped for new expressive possibilities in vocality and communication.

In music, speech synthesis is treated with the same expressive curiosity as the sung voice or another instrument. Voder technology, once used to mask telephones from eavesdroppers during World War II, was repurposed as the vocoder, which has been creatively explored and used extensively in music (Tompkins, 2011).¹⁴⁰ As Eshun (1998) describes in his book *More Brilliant than the Sun: Adventures in Sonic Fiction*:

The vocoder turns the voice into a synthesizer. Electro crosses the threshold of synthetic vocalization, breaks out into the new spectrum of vocal synthesis. It synthesizes the voice into Voltage, into an electrophonic charge that gets directly on your nerves. Turning the voice into a synthetic spectrum of perverse voco-imps lets you talk with cartoons, become cartoon, become animal, become supercomputer (06[080]).

Eshun’s description, and the understanding, of voder and vocoder speech synthesising technologies in the musical field, shows a vast distinction from the perspective of conversational AI creators, as described earlier by Oord et al. (2016). Eshun observes the transformative nature of voice synthesis to become other real or imagined beings, non-human animals or inanimate matter. By appreciating synthesised voices as material, they can morph and shape-shift according to other frameworks of understanding that do not have to map onto an envisioned human individual or prescriptive profiling practices.

Making of Acoustic Ecology of an AI System and Technical Considerations

¹³⁹ Beginning at 0:52s of the video recording.

¹⁴⁰ Dave Tompkins writes extensively about the history of electronic voices and the history of the vocoder (Tompkins, 2011)

'Acoustic Ecology of an AI System' utilises Kate Crawford and Vladan Joler's (2018) work *Anatomy of an AI System*, which investigates the deep material networks of an Amazon Echo device as an anatomical map. I imagined what each location on Crawford and Joler's map would be like to sound my voice within. I then designed acoustics to explore using sound to locate disembodied, synthesised voices in space, time, environment and architecture, in order to reveal and resist vocal profiling. These sonic environments contextualise the voice within a broader ecological and embodied system that underpins a conversational AI system, evidencing that a voice never acts in isolation, but is always co-created in its sounding with other voices, bodies, environments and matter. The seven points from the map I chose to create into sonic environments were 'Mines', 'Smelters and Refiners', 'Component Manufacture', 'Assemblers', 'Transportation', 'Data Labelling' and 'AI Training' as described by Crawford and Joler. Experimenting with the sonic design of AI's synthesised voices could provide a means to conceptualise and navigate the unseen networks of AI systems to represent them more holistically, therefore, adding a sense of the material world that fundamentally encompasses these digital systems.

Digitally manipulated audio environments were created to produce acoustics for the seven environments chosen for 'Acoustic Ecology of an AI System'. The audio plug-ins RaySpace and Crowd Chamber, created by company QuikQuak, were used in conjunction with Audacity audio editing software to achieve this.¹⁴¹ RaySpace, a 'room simulator', was chosen over others because I was able to visually draw spaces, including their form, height, width and internal features, which is not usually possible in similar software packages (Figure 40). Crowd Chamber can simulate small to very large crowds of voices and vary their spectral content and delay to achieve different chorusing effects (QuikQuak, n.d.) (Figure 41). Electronic music producer Sam Kidel created a related project in his track *Live At Google Data Centre*

¹⁴¹ Audacity is a free, open-source easy-to-use, multi-track audio editor and recorder for Windows, macOS, GNU/Linux and other operating systems (Audacity, n.d.).

(2018):

In a process he describes as “mimetic hacking”, Kidel uses architectural plans based on photos of the data centre to acoustically model the sonic qualities of the space. The resulting acoustics on *Live at Google Data Center* simulates the sound of Kidel’s algorithmically-generated notes, rhythms and melodies reverberating through the space, as though a bold illegal party was being held in the maximum security location (Opiah, 2018).

Kidel’s depiction of an illegal rave, acoustically situated within Google’s Data Centre, presents the fantasy of occupying a space that is out of bounds to most people yet contains digital architecture that governs and ‘looks in’ on many aspects of our lives. Kidel also uses acoustic design to transport people to locations we cannot physically see or fully comprehend. His piece is both a critique and a creative exploration of sound design, as similarly explored in ‘Acoustic Ecology of an AI System’.



Figure 40: Ray Space audio plug-in software screenshot. (QuikQuak, n.d.).



Figure 41: Crowd Chamber audio plug-in software screenshot. (QuikQuak, n.d.).

RaySpace and Crowd Chamber allowed me to work with the sound and sounding of synthesised voices as material. Instead of experiencing these voices purely as disembodied and acousmatic, as they are currently presented in conversational AI systems, I wanted to situate them within environments, space and

time so that they could be located and heard to narrate the conversational AI system itself through designed acoustics and sound. The process of Speculatively Voicing and designing the sound I desired for each clip was a process of trial and error. I would make minor adjustments to the audio, listen, adjust again, listen again...until I found a sonic effect for each voice, which I felt represented the environment I was trying to highlight from *Anatomy of an AI System* (Crawford & Joler, 2018). Sterne notes the intuitive nature of designing sound to emulate specific environments or effects. For example, the “cathedral” setting on a reverb device bears that name because it sounds like a cathedral to the designer, not because it has any actual relation to any particular cathedral’ (Sterne, 2015, p. 123). This description by Sterne reflects the way I have attempted to design sound in this project. In the table in Figure 42, I note the qualities of sound I was hoping to achieve with the final sounded audio clips ([Item 24](#)) in order to represent these seven different environments or situations.

Clip No.	Audio Title	Location on Map	Voice Sound and Sounding Qualities
1	Mining	Mines	Long echo. Cold. Large expansive space. Wet. Damp. Metallic.
2	Smelting	Smelters and Refiner	Hot. Melting. Dripping. Vaporous. Shimmering.
3	Industrial Manufacture	Component Manufacturers	Background noise. Distant voices. Industrial. Voices engulfed by machinic noise.
4	Assembly Line	Assemblers	Repetitive. Uniform. Close. Detailed.
5	Cargo Boat Distribution	Transportation	Hollow. Boxy. Metal container. Confined on all sides.
6	Data Processing	AVS	Digitised. Detailed articulation. Magnified voice sounds.
7	AI Training	AI Training	Long drawn out voice. Stretched. Pulled in many directions. Expansive. Movement.

Figure 42: Acoustic Ecology of an AI System table of Audio Title, Locations on Map, Voice Sound and Sounding Qualities table.

Google Duplex is an AI-enabled assistant with a synthesised voice created to sound as natural as possible and designed to emulate human communication, complete with prosody, pauses and punctuation (See: Leviathan & Matias, 2018). This application was created using Google's WaveNet, a deep neural network for generating raw audio, which is trained with recordings of real speech and has enabled the creation of relatively realistic-sounding human-like voices. However, when training the network without the text sequence, it still generates speech, but now it must make up what to say (Oord & Dieleman, 2016). 'WaveNet Babble' sounds like speech, but its linguistic content is void.¹⁴² Although nonsensical in a

¹⁴² The 'WaveNet Babble' audio clips are freely available to download from the Google Deepmind Blog (See: Oord & Dieleman, 2016).

linguistic sense, it provides a representation of what synthesised voices in conversational systems currently sound like and exposes AI's process of training synthesised voices. It is used as source audio for this project and is appropriate as it concerns only the sound and sounding of speech, investigating the sonic potential of voice. As a designer and singer investigating voices in conversational AI systems, I aim to advance an understanding of the voice as sonic material that can be used for speculative design research and practice. With the growing use of AI voice technologies, it is crucial to investigate how vocal sonic material can be manipulated and designed to represent a multitude of situations and scenarios.

'Acoustic Ecology of an AI System' was presented on the Research and Waves platform as an online interactive audio experience (Abbas-Nazari, 2020). Visitors to the website are invited to 'drag the white dot across the screen to explore the sounds within the black box'¹⁴³ (Figure 43). Minimal visual simulation and no specific visual indicators were provided to locate the seven separate audio clips positioned within the 2D screen space as a way to focus participants' listening faculties to wander through the audio landscape ([Item 25](#)). Participants must actively use their imagination to conjure the 3D space, potentially filling in visual cues or simply listening and conceptualising, embodying the fictional environment. Each audio clip sonically describes and aurally illustrates a different 3D environment animated by a synthesised speaking voice. To add a further sense of spatiality, as people move the white dot around the screen, the audio clips fade in and out to blend and blur the different clips together.¹⁴⁴ The audio experience designed for 'Acoustic Ecology of an AI System' could be compared to virtual reality but because it is purely audio-based, it can perhaps be described as a 'virtual audio reality'.

¹⁴³ Thanks to Henrik Nieratschker from the Research and Waves Collective for inviting me to participate and also for his skills and expertise in helping to configure this project for an online audience.

¹⁴⁴ This borrows from a technique often used in computer gaming where sound designers will 'mix' audio from different scenes as players move through different gaming environments to give a sense of moving through 3D or real-life environments. Thanks to Royal College of Art extended reality (XR) technician Thomas Deacon for recommending this technique and describing how to achieve it.

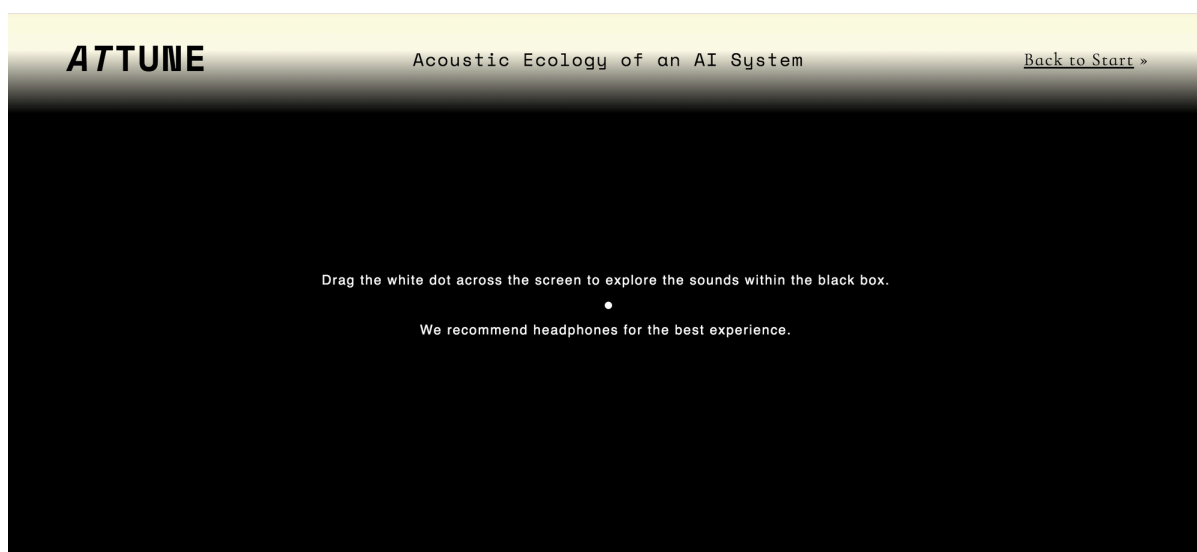


Figure 43: Screenshot of home page of Acoustic Ecology of an AI System on Attune / Research and Waves website. (Abbas-Nazari, 2020).

Speculatively Voicing Synthesised Voices in Conversational AI Systems

As previously described in this thesis, the four conditions of materiality, grouped under the Speculative Voicing Framework, are as follows:

One voice in conversational AI systems can:

1. be embodied and co-created with other matter and/or the environment
2. be embodied and co-created with many other bodies and voices, such as in a choir
3. embody many voices, co-created with the body
4. embody many disembodied voices, co-created with the conversational AI system

I employ these four conditions to assess the ability of my case study project, 'Acoustic Ecology of an AI System', to produce vocal imaginaries with polyphonic potential. In the following section, I describe why synthesised voices in conversational AI systems should be comprehended in this manner, moving towards an intersectional appreciation of the voice with an ecological and social sensitivity. The forthcoming sections of text respond to answer Q2 & Q3 of my research: How

can applying this [Speculative Voicing] methodology reveal and resist vocal profiling, in the AI era?

One voice in conversational AI systems can:

1. be embodied and co-created with other matter and/or the environment

While AI is often perceived and understood as automatic and enabled by computing power, the extensive reliance on human labour, both physical and cognitive, is hidden and unrepresented in these systems and our interactions with them. In order to dive deeper into understanding the materiality of conversational AI systems, research compiled by Kate Crawford and Vladan Joler provides an investigation of the deep material networks of an Amazon Echo device, voiced by Alexa, in their project *Anatomy of an AI System* (2018). As Crawford says, 'Put simply: each small moment of convenience – be it answering a question, turning on a light, or playing a song – requires a vast planetary network, fuelled by the extraction of non-renewable materials, labor, and data'. Their research reveals that networks of conversational AI systems are immaterial and intangible but equally materially embedded in the geographical and geological landscape. On the left-hand side of their 'anatomical' map, Crawford and Joler draw attention to the environments and locations where different aspects of material production for an Amazon Echo device occur.¹⁴⁵ In his book *The Stack* (2016), Benjamin Bratton says that 'we carry small pieces of Africa in our pockets'. What he's alluding to is the fact that one computer chip is composed of over 60 elements, mined from locations far removed from our own physicality but which are now embedded in our daily lives. This sentiment is equally true of a conversational AI device such as Echo. All its constituent parts may now be contained within one plastic casing, sitting in a family kitchen, but they were originally derived from many different countries. 'Acoustic Ecology of an AI System'

¹⁴⁵ The maps' locations are titled 'Mines, Smelters & Refiners, Component Manufactures, Transportation, Assemblers and Distributors'. Running adjacent to this are the aspects of digital production and their environments, headed: 'Data Preparation and Labelling, AI Training, Amazon Inc. Infrastructure, Internet Platforms & Services, Internet Infrastructure and Domestic Infrastructure' (Crawford & Joler, 2018).

seeks to sonically illustrate the divergent spatiality and temporality occupied by these manufacturing processes, which are continually in motion and enable the functioning of these devices. I acoustically design synthesised voices to situate them in these various locations to imagine alternative ways to present these voices, that do not rely on normative expectations or practices of profiling. Moreover, this approach provides a means to conceptualise deep and distributed networks of conversational AI systems and prompt critical conversations about their environmental impact.

One voice in conversational AI systems can:

2. be embodied and co-created with many other bodies and voices, such as in a choir

Particularly fascinating, yet profoundly worrying, is the people hidden within Amazon's conversational AI system. Twenty-nine instances of human labour are found on Crawford and Joler's schematic. Some of these people are referred to as mechanical Turks¹⁴⁶ or crowd workers – people hired by businesses to perform on-demand tasks remotely, which computers cannot fulfil. Employers advertise jobs known as Human Intelligence Tasks (HITs), and human labourers complete work such as specifying content in images or video, writing product descriptions, or answering questions (mturk, n.d.). In other words, tasks that the devices' users expect the technology, or AI, to fulfil are actually performed by humans. Here, we can understand that a synthesised voice, in this instance, does not and should not signify one being, body or entity. It is, in fact, a chorus of workers contributing and collaborating to express the singular-sounding voice of a conversational AI system. Many human labourers in conversational AI systems work in precarious and/or hazardous conditions – in mines, for example, where dangerous gasses and fumes exude into the environment. The human labourers are not directly represented in

¹⁴⁶ The original 'mechanical turk' was a fraudulent chess-playing device that people thought to be fully automated, created in 1769. Audiences were convinced it was an 'artificially intelligent' machine. However, it secretly concealed a chess master playing the game (mturk, n.d.). It was designed by Wolfgang von Kempelen, who also made mechanical speech synthesis machines around the same time (See Introduction).

‘Acoustic Ecology of an AI System’; instead, a choir of synthesised voices animate locations where this labour occurs. It is intended that listeners of the project start to imagine what it might be like working in these environments while contemplating the extensive physical, cognitive and digital processes performed by humans.

One voice in conversational AI systems can:

3. embody many voices, co-created with the body

Being or presenting as multiple is already a strategy used when sounding as a singular voice is highly restricted and can cause harm, should it become apparent. An instance of this is in the Islamic Republic of Iran, where female singers are banned from singing alone. As described by Yara Elmjouie, performers employ ‘*ham-khani*’, or ‘co-singing’, in which solo female vocalists sing with other men and women to mask their own voice to negotiate and circumvent these rules. Sometimes this co-singing is similar to choral singing, but at other times the additional performers have been known to quietly hum along or even mouth the lyrics while the soloist sings (Elmjouie, 2014). This distributed embodiment of a co-created voice acts as a safety-in-numbers scheme. It is utilised in ‘Acoustic Ecology of an AI System’ to create depictions of synthesised voices as multiple. Sometimes the voices exhibited are ambiguous in terms of how many are present, but also in that seven different versions of voice(s) are presented. This, too, can be seen as a form of protection, resisting current vocal profiling practices that entail depicting women in derogatory ways.

One voice in conversational AI systems can:

4. embody many disembodied voices, co-created with the conversational AI system

Synthesised voices in conversational AI occupy a non-binary space, simultaneously maintaining conditions of being human and non-human. They are trained on human voice data, created and maintained by humans (Crawford & Joler,

2018) and aim to be understood as human. Nevertheless, they are non-human, synthetic renderings of voice which currently portray very narrow representations of what it is to be human (West, Kraut & Chew, 2019). Being digitally created, there is no reason for synthesised voices to be represented and understood as singular. This is especially true, in light of their underpinning by large groups of unrecognised human labour with vast geological distribution. This research highlights the sound and sounding of voices as multiple to explore how voices are always polyphonic and can never be singular. 'Acoustic Ecology of an AI System' rejects the idea and understanding that the voice can or must signify and represent an individual person, as currently prescribed by normative expectations in vocal profiling in conversational AI systems. By Speculatively Voicing synthesised voices, I seek to exploit vocal materiality and the concept of polyphony to illustrate how voice in conversational AI occupies multiple states and disembodiments simultaneously.

Analysis of Acoustic Ecology of an AI System

My autoethnographic, thick-description analysis¹⁴⁷ of this case study practice project aims to highlight the divergence between current voice-profiled presentations of synthesised voices in conversational AI systems compared to the exploration and potential sounding of voices in this research, defined by the experimental methodology. Contrasting the two modes of understanding the sounding of voices builds critique by revealing and resisting vocal profiling, addressing questions 2 and 3 of my research. In order to do this I will use two audio clips as exemplars: speaker-2 ([Item 26](#)), the original unedited audio of Google WaveNet Babble (Oord & Dieleman, 2016), and Clip 2 'smelting' ([Item 27](#)), one of the manipulated audio clips from 'Acoustic Ecology of an AI System'.

WaveNet Babble speaker-2 is a female-sounding synthesised voice. It cannot be visually observed, only heard when its sonic force vibrationally encounters the ear's

¹⁴⁷ See Chapter 4: Methodology for further information.

tympanic membrane. The nonsensical voice's physical or geographic origin cannot be discerned via spoken words, language or audible sounding. This voice is static in space and acoustically restricted. Sonically, it for me resonates with the speech of Scandinavian friends and family, with their bright, extended vowel sounds produced by the raised cheeks of a widely opened mouth, with teeth showing. In my auditory imagination¹⁴⁸ I summon the physiology of a face, its interior, and its movements based on my experience as a singer and singing in various languages. I can only speak English, but I've learned to reproduce foreign speech sounds through my choral singing practice, by listening and then morphing my orality to mimic what I hear. In speaker-2 I sonically perceive the tongue, occasionally pressing the back of the top row of teeth, funnelling the sound through a squeezed throat, meeting the intersection of teeth and tongue, reverberating between wet fibrous flesh and smooth enamel. Sometimes the front of the tongue is behind the lower teeth, and the mid-tongue is arched towards the roof and front of the mouth. Air is forced out the corners of the lips, producing a slight hiss. Again, the intonation of the phrase reminds me of my Iranian-Swedish cousin talking to her children in a melodic, almost song-like, calming temperament.

Audio Clip 2, 'smelting', is a speculative voicing. There is the presence of the synthesised voice (originally speaker-2); however, equally, there is the existence of an environment that envelops the oral entity. The voice occupies and is situated in a space, and the space shapes the vocality. The sounded voice and environment are co-created in reciprocal materiality, in which the sound produced is not indicative of a singular individual or being but evocative of an entire setting or scenario. This voice is an ecology of vaporous, liquid, hot, vibrational energy. The environment is cavernous but consuming. It refuses to contain the voice, sustaining its thick, muggy sonic residue long after its initial indentation on the air. The voice being produced

¹⁴⁸ A term used by Voegelin (2014) who writes extensively on sound and imagination, especially to produce 'possible worlds'.

could be emerging from many bodies or the sonic temporality of the space itself. At times it seems to be melting into air, merging into an unseen but acoustically deduced architecture. I have never been to a smelting furnace where the extraction of rare earth metals takes place,¹⁴⁹ but I can speculate on and imagine its sensorial and material characteristics. I illustrate this through designed sound based partly based on previous singing experiences in different environments and architecture. I endeavour to encapsulate the material qualities of smelting metal through the design and sounding of voice.

Conclusion

Synthesised voices are by nature disembodied, as I have described. However, this thesis and its methodology explore voices from their embodied origins. Researchers at the University of Florida (Blue et al., 2022) analysed deepfake AI-synthesised voices¹⁵⁰ through simulated vocal tract reconstruction. They found that they produce vocal tract shapes that do not exist in people. Rather than being contoured and organic in shape, the recreated vocal tracts resembled ‘the size and shape of a drinking straw’ (pp. 2702-2703). This suggests that the ‘embodied’ voices of synthesised voices also exist as though they are in a materially detached vacuum. While the research by Blue et al. (2022) shows that synthesised voices cannot ever be returned to an embodied state, the approach of ‘Acoustic Ecology of an AI system’ allows for the synthesised voices of conversational AI system and their (hidden) bodies, embedded with the material world, to be more conscientiously recognised.

Anatomy of an AI System is used as a template, or map, to stipulate the seven audio clips and how the acoustic spaces were modelled and defined, due to its detailed account of the material and embodied nature of these systems. The design

¹⁴⁹ This is the basis of the location I acoustically designed for Clip 2 ‘smelting’, referencing research by Kate Crawford and Vladan Joler in their project *Anatomy of an AI System* (2018).

¹⁵⁰ A deepfake voice aims to closely mimic a real, known person. Although synthesised using AI processes, they aim to ‘accurately replicate tonality, accents, cadence, and other unique characteristics’ (Weitzman, 2022).

process was partly based on available information and partly imagined – in other words, it was a speculative venture. For example, Crawford and Jolar document that part of Echo's manufacturing takes place in mines where human labourers extract rare earth minerals. For this audio clip, a large acoustic environment was digitally modelled, which created a very 'wet' acoustic, which was applied to the source audio to imply a cavernous space. On reflection, I do not know that this is what the environment is really like, but it still aims to transport the listener to a cave or mine-like space sonically. Other elements of the assembly process are even more sonically ambiguous, especially as the processes become less physical and more digitally focused. For example, what does an AI training process sound like? There may be no way to know, but we can use sound to unlock our imagination and locate ourselves in that space. These processes may be digitally formulated, but they are still materially co-created. Even though we may not humanly be able to hear or perceive the sound, sonics are still being created.

He (2019) says we could avoid the 'uncanny valley' (Mori, MacDorman & Kageki, 2012) altogether and preserve recognisably robotic speech-to-text voices as an important artistic aesthetic even as speech synthesis technology advances, in order to be able to distinguish between human and non-human entities. However, as 'Acoustic Ecology of an AI System' shows, Speculative Voicing presents an alternative framework to design synthesised voices. This framework transcends categories of self/other, human/non-human, working with vocal materiality to erode these binary distinctions and shift the discussion to encompass more than just gender bias and surveillance enabled by machine listening.

This chapter completes the account of my case study practice works. In the following chapter, I evaluate my two case study projects through a workshop with industry professionals who design and work with conversational AI systems.

Chapter 7: Evaluation of Speculative Voicing

Introduction

This research, led by practice, has evolved a new materialist, intersectional sonic speculative design methodology. This was then applied in order to reveal and resist vocal profiling: specifically, the AI profiling of sounded human voices to predict faces and the vocal sounding of synthesised voices, in conversational AI. Since this research aims to critique current vocal profiling practices in the field of AI, it is necessary to evaluate my methodological propositions within an appropriate context.

Workshop Structure and Process

I conducted a workshop with employees from IBM, based in London.¹⁵¹ IBM was identified as a company to engage with in my evaluative process primarily because my PhD studentship, supported by TECHNE and their National Productivity Investment Fund (NPIF), initiated a pre-arranged partnership with the company (TECHNE, n.d). IBM is a valuable organisation to comment on research involved with AI as they offer an AI-based ‘portfolio of business-ready tools, applications and solutions, designed to reduce the costs and hurdles of AI adoption while optimizing outcomes and responsible use of AI’ (IBM Watson, 2021). These services are grouped under the umbrella name ‘IBM Watson’, after IBM’s founder, Thomas J. Watson. IBM was also involved in early speech recognition experiments, with the IBM Shoebox device designed in 1961, and in the same year, programmed a computer to sing *Daisy Bell* (Radovic, 2008). Therefore, it is appropriate to work with IBM, as they not only actively work with cutting-edge developments in AI but were also involved with the early foundations of what defines conversational AI systems today.

¹⁵¹ IBM (International Business Machines) is a multinational technology corporation operating in over 171 countries.

The two-hour online workshop schedule (Figure 44) included leading participants through simple Deep Listening exercises by Pauline Oliveros: *Imaginary Meditations* (1979), *Ear Piece* (1971) and *Your Voice* (1974), to encourage consideration of the materiality of voice and sonic thinking.¹⁵² We briefly discussed Oliveros' ideas of hearing versus listening (Oliveros, 2015). We endeavoured, as Oliveros says, to 'expand your receptivity to the field of sound by defocusing your ears as you would your eyes for a wider visual field', from her piece *All or Nothing* (Oliveros, 2013). This 'warm-up' for the main workshop exercise was intended to guide participants into actively thinking through sound. As with the analysis method¹⁵³ I myself used in this research, I encouraged participants to focus on the synaesthetic, multi-modal qualities that sound possesses (Van Leeuwen, 2016) and use descriptive words to give an account of what they had heard. I asked participants to respond in this fashion to 18 audio clips in total. The first seven audio clips were those created for 'Acoustic Ecology of an AI System' ([Item 24](#)) and one additional 'control' audio sample of unedited, unprocessed Google Wavenet Babble¹⁵⁴ ([Item 28](#)), for later comparison, numbered 1-8. The remaining 10 audio clips, labelled A-J were extracted from the project 'Polyphonic Embodiment(s)' ([Item 21](#)). Participants noted their descriptions in the collaborative online digital workspace application Mural (Appendix C / [Item 29](#)). The participants were given no particular indication or information about what they were listening to, to avoid influencing their impressions of what they heard. Later in the workshop, I gave an overview of my PhD research via a presentation documenting my methods and methodologies, key themes, and outlines of my practice projects. This was presented after the main exercises to contextualise the origins and intentions of the audio clips they had listened to previously. It was also an opportunity to initiate a discussion of key themes in my research relating to vocal profiling in conversational AI systems.

¹⁵² These three pieces by Oliveros were chosen because they can be performed easily and individually – by simply reading the written text – which suited an online workshop setting.

¹⁵³ See Introduction for more specific information on this method.

¹⁵⁴ See Chapter 6 for greater explanation.

Time	Duration	Activity
14:00 - 14:10	10	Introductions. Discussion of consent and ethics approval of the workshop
14:10 - 14:20	10	Introduction to workshop with short presentation of examples from <i>Speculative Listening</i> workshops
14:20 - 14:30	10	Exploring different ways to listen. Deep Listening exercises
14:30 - 14:40	10	Introduction to Mural and practice example
14:40 - 14:55	15	Listen to different audio clips of designed voices prepared by the researcher (<i>Acoustic Ecology of an AI system</i>). Participants will respond to what they hear using descriptive words and / or drawn images
14:55 - 15:10	15	Listen to different audio clips of designed voices prepared by the researcher (<i>Polyphonic Embodiment(s)</i>). Participants will respond to what they hear using descriptive words and / or drawn images
15:10 - 15:20	10	Discussion, reflection and feedback from participants on prepared activities
15:10 - 15:30	10	BREAK
15:30 - 15:45	15	Presentation by researcher on the wider context of the projects that the audio clips derived from and PhD research themes being explored.
15:45 - 15:55	10	Open discussion, reflection and feedback from participants on themes of PhD research
15:55 - 16:00	5	Complete Google Form questionnaire. Goodbyes.

Figure 44: IBM Workshop Schedule.

Workshop Participation and Participants

All four participants were employees at IBM's London headquarters and included:

An AI senior management consultant who has been 'designing and delivering AI solutions for the past 7 years'.

A data scientist in AI and analytics with a background in linguistics and a PhD in conversation analysis, who stated that 'virtual assistants has really been my main focus at IBM delivering, deploying and designing journeys for virtual assistants with our clients across many different industries'.

A recently appointed employee of IBM working in AI and analytics, who said

they are positioning themselves ‘now more towards natural language processing because I think it's really exciting’.

An AI consultant who works in conversational AI, with a background in linguistics and a Master's in speech-language processing, working on projects with chatbots and virtual assistants, interactive voice response (IVR), telephony and conversational design.

Participants were self-selecting via an email ‘call-out’ to employees through my main point of contact at IBM. In the call-out, I provided a schedule for the workshop and the Participant Project Information & Consent Form. I explained that all were welcome to join, but the workshop might best suit those interested in sound, sound design, voice user interfaces, interaction design, user experience, user interaction, conversational AI and persona design. These roles were highlighted to capture the views and perspectives of those at IBM who are most relevant and valuable to evaluate the effectiveness of the research. Three of the four participants participated fully in the workshop, and one engaged partially in the first evaluative exercise.

I evaluated the feedback from the IBM participants against the four conditions of the Speculative Voicing Framework I defined to explore voice profiling in conversational AI systems. This was to examine whether the case study projects I produced exhibited the polyphonic features of voice I aim to bring to the fore. All quoted information from the participants in this chapter is extracted from a transcript of the workshop.¹⁵⁵

Evaluation of Polyphonic Embodiments

The audio clips labelled A-J were produced as part of the project ‘Polyphonic

¹⁵⁵ The full transcript has been omitted as part of the thesis submission for reasons of research ethics. If you require further information please contact the author.

Embodiments(s)' and were designed and intended to be 'listened' to by an AI.¹⁵⁶ However, they were presented during the IBM workshop as part of this PhD evaluation to gain a greater understanding of their polyphonic potential, using the Speculative Voicing Framework, in which:

One voice in conversational AI systems can:

1. be embodied and co-created with other matter and/or the environment

I questioned participants about this set of clips, asking if they thought they were voiced by a human or were synthesised. One participant said, 'they felt a bit more human because they had this feature that, that felt distinct to owning a face and a mouth, as opposed to just text-to-speech'. I found it interesting that the participant determined that the voices were not just TTS because they could hear the presence of bodily features, which synthesised voices do not possess. The participant also, comprehended the embodied state of the voice. Participants were frequently observed describing facial features, documented in Mural (Appendix C/ [Item 29](#)), particularly the mouth, which they sometimes understood to have been interfered with. For example: 'seems like they have an object in their mouth and between their teeth blocking the tongue'. When asked to describe what they heard in this set of clips, generally, a participant said, 'I felt like they were playing with their faces in order to make sounds and I felt they used the objects to do that'.

One voice in conversational AI systems can:

2. be embodied and co-created with many other bodies and voices, such as in a choir

At no point did participants reference hearing multiple voices in any of the audio clips. Due to the nature of this practice project and the way it was presented in the workshop, this condition of the Speculative Voicing Framework was difficult to achieve in this instance. However, following the exercise I showed participants video

¹⁵⁶ See Chapter 5 for more information about this project.

documentation of the 'Polyphonic Embodiment(s)' project ([Item 22](#)), in which I was filmed performing the different voiced audio. It transpired that they had not realised it was my voice they had heard, despite conversing with me for around 40 minutes before hearing the clips. In addition, participants alluded to multiple 'bodies and voices' – see below.

One voice in conversational AI systems can:

3. embody many voices, co-created with the body

Audio clips labelled A-J, extracted from the 'Polyphonic Embodiment(s)' project were all performed by me. In 18 out of a possible 30 instances, participants referenced a person. However, there was some variation in the person they reported. One participant mainly described sonic features they heard but did use 'they' on one occasion. Another participant noted 'female', 'speaker', and 'emulating an old person'. The third participant described 'young boy', 'a teacher', 'boy', 'speaker', 'person', 'she', and 'woman'. In evaluating this project, I realised that my embodied voice had embodied multiple voices and bodies. In turn, the audio produced for the 'Polyphonic Embodiments(s)' project was considered polyphonic, exhibiting the potential to be understood as multiple voices by human listeners. This points towards common misconceptions of the ability of human voices to be profiled by both humans and AI-enabled systems within current frameworks in which one body can be attributed to one voice and vice versa.

One voice in conversational AI systems can:

4. embody many disembodied voices, co-created with the conversational AI system

A distinction lies between human listeners and AI in response to the 'Polyphonic Embodiment(s)' audio clips. The wav2face AI was observed generating a whole new face for each DIY voice-manipulated clip, as shown in Chapter 5. However, during this workshop human participants could detect that facial features had been altered

to manipulate the voice they heard. For example, in clip A, ‘the sound seems bothered by the person having their hand over their mouth’, and in clip J, ‘spoken through a cardboard tube’. As such, conversational AI systems appear to exaggerate and embellish disembodied voices, a phenomenon also supported by the theoretical research in this PhD.¹⁵⁷

Evaluation of Acoustic Ecology of an AI System

The audio clips labelled 1-7 were produced as part of ‘Acoustic Ecology of an AI System’. Audio clip 8 acted as a ‘control’ example, representing how voices currently sound and are sounded in conversational AI systems. I presented clips 1-8 during the workshop for IBM employees as part of this PhD evaluation to gain a greater understanding of the polyphonic potential of the work using the four conditions in which:

One voice in conversational AI systems can:

1. be embodied and co-created with other matter and/or the environment

For the un-manipulated, unedited ‘control’ audio clip 8, interestingly, all four workshop participants described physical or personal attributes of a person. This included ‘sounds like speaker has a stuffy nose’, ‘confident voice [...] reminds me of Scottish accent’, ‘someone speaking (really) fast’, and ‘dominant female voice’. In the remaining clips, however, on only four occasions, out of a possible 28,¹⁵⁸ did participants describe physical or personal attributes of a person. In clip 4, for instance ‘person seemed sad’. This feedback is congruent with my expectations that listeners try to deduce a person from the sounded voice in current conversational AI systems. The previous seven audio clips from ‘Acoustic Ecology of an AI System’ followed the methodology defined by this thesis research. Participants were more inclined to describe a location or atmosphere with these audio clips. While

¹⁵⁷ See Chapter 1

¹⁵⁸ Four participants multiplied by seven prepared audio clips = 28.

sometimes participants would note the gender of the voice they heard, they would situate it in an imagined scenario or setting. For example, clip 7 'sounds like a dome or someone in a church', clip 5 'a conference with a badly set-up sound system', clip 1 'feels like they are outside', and clip 3 'reminds me of feeling being in an airport'. Participants also commented on how the heard audio made them feel when they heard it. For example, clip 1 'disconcerting – feel a bit uncomfortable hearing it'. Overall, participants were far more inclined to describe sonic rather than personal features.

One voice in conversational AI systems can:

2. be embodied and co-created with many other bodies and voices, such as in a choir

In two of the seven audio clips, participants noted that they heard multiple voices. For example, clip 4, the listener notes 'feels like 3 different people speaking', and clip 7 'a group of people responding in unison'. Alternatively, participants alluded to multiple people being present in the environment, despite the fact that they did not actually speak: clip 3, 'someone making an announcement to many people', and clip 5, 'sounds like a guide thanking someone for being there'.

One voice in conversational AI systems can:

3. embody many voices, co-created with the body

The findings from this evaluative exercise confirmed that current frameworks for the sound design and sounding of synthesised voices in conversational AI systems (represented by clip 8) encourage listeners to pinpoint physical or personal attributes of a human. This serves as a critical point of reference for vocal profiling, also reinforcing the validity of profiling. Meanwhile, in the other seven prepared audio clips, participants validated the intentions of this Speculative Voicing enquiry as they described events, environments, and settings and attributed the sounding to multiple voices or people. Speculative Voicing, as a renewed methodology for the

sound and sounding of voices in conversational AI systems, could provide a means to resist and disrupt profiling practices. Nevertheless, the participants, for all but one of the clips (clip 2), were observed to continue to reference in some way a ‘voice’, ‘speaker’, ‘person’ or ‘people’. This methodology suggests, therefore, that Speculative Voicing could be used as an alternative to vocal profiling without becoming purely sound or noise.

One voice in conversational AI systems can:

4. embody many disembodied voices, co-created with the conversational AI system

Participants did not describe the sonically illustrated environments as I had designed and intended. For example, clip 3 was described as ‘synthetic low wind [...] dystopian’, ‘an airport’, ‘a rocket ship’ and ‘waves’. However, the audio was intended to evoke industry, manufacturing, and/or a large factory atmosphere.¹⁵⁹ When surveyed, participants expressed quite varying opinions of what they had heard from all the clips. Participants were not provided with information about the project’s motivations and intentions before the exercise, to avoid influencing their responses. Owing to the ambiguity that sound can afford if narrative prompts were made available, it is more likely that participants would identify the designed sonic environments with more clarity. For the purposes of this PhD research and its aims, participants did not need to reach a consensus on the listening materials provided. This evaluation tested the polyphonic potential for vocal imaginaries to reveal and resist vocal profiling via the Speculative Voicing methodology, which was achieved in this instance.

Analysis of Findings

The table below (Figure 45), summarises the core thematic features of voice, identified in this research, of vocal profiling and of the Speculative Voicing

¹⁵⁹ See Chapter 5

Framework, under the headings 'A' and 'B', respectively.¹⁶⁰ Examples of evidence from the workshop participants' Mural responses is cited in the third column, supporting the qualities of Speculative Voicing being achieved in the two case study practice projects.

A	B	B
Profiling	Speculative Voicing	Example Evidence
Visual	Sonic	<i>Clip 6</i> : "can hear the mouth full of the woman, as if she has too much saliva while speaking,"
Singular	Multiple	<i>Clip 4</i> : "feels like 3 different people speaking"
Fixed	Polyphonic	<i>Clip H</i> : "sounds a bit like voice is trying to emulate an old person (maybe with missing teeth?)"
Stable	Co-created	<i>Clip 7</i> : "sounds like a dome or someone in a church chanting what the person is saying."
Objective	Subjective	<i>Clip 7</i> : "reminds me of threatening scenes in thrillers."
Knowable	Malleable	<i>Clip E</i> : "sounds like speaker is speaking around something in their mouth"
In a Vacuum	In Space and Time	<i>Clip 1</i> : "far away, looong, tired, last part feels like they are outside"
Binary	Non-Binary	<i>Clip 3</i> : "sounds like a rocket ship launch announcement, someone making an important announcement to many people "
Normative	Explorative	<i>Clip C</i> : "the mouth of this young boy seems closed making an oval shape with their mouths, i feel their eyes are also slightly closed"

Figure 45: Example evidence of Speculative Voicing methodology from IBM Workshop evaluation.

The emergent themes recapitulated in the table (Figure 45) describe the voice

¹⁶⁰ This A/B table, in part, borrows from Dunne and Raby's A/B, 'Affirmative / Critical' table, which also functions as a manifesto for speculative and critical design practice (See: Dunne & Raby, 2014, p. vi-vii).

enacted within two different ontological and epistemological rationales. The summarised findings establish how the use of the Speculative Voicing framework, as a sonic speculative design methodology, provides an alternative format to comprehend voices in AI. The Speculative Voicing framework, which reveals and resists profiling, provides an awareness of voicing to suppress normativity and marginalisation enabled within profiling. The contrasting A/B positions can be used to inform and advocate for more responsible and conscientious development in conversational AI systems.

Workshop Discussion and Wider Implications

During open discussion in-between the formal exercises, participants confirmed and reinforced that in AI, profiling imaginary human people is standard practice when defining and designing synthesised voice and speech, in which voice and speech are directly linked to facets of identity and personality. One participant commented that when designing personas for virtual assistants, they encourage clients to consider the following: ‘what does the person look like or [...] where did they go to school. I try to really humanise the virtual assistant as much as possible’. This confirms ‘vocal profiling’ as a useful term, and relevant concept to prompt interdisciplinary discussion about the sound and sounding of voices, distinct from terms such as ‘biometrics’, which deals more specifically with data about humans. IBM participants also described that with other clients, little or no thought is given to the voice of AI-enabled assistants – one participant described how, with a client, ‘one of their employees [...] used to be a radio presenter and they took his voice because, because err, that was the kind of most convenient. I don't know if they thought a lot about, kind of, voice and how it matched the brand or anything’. The comments above indicate that while profiling is the dominant method used to determine the sound and sounding of synthesised voices in the field of AI, more broadly, those working in other industries lack the knowledge to make informed decisions and understand what the sound of voices could potentially implicate when utilised for

different conversational AI products or services.

One unsettling aspect of the workshop discussion led participants to consider how ideas around Speculative Voicing and its framework, which I presented, could create more personalisation of, or for, synthesised voices. The participant elaborated: 'the client asked me can we, based on the person's rank, can we give them a different type of answer?' They added that, 'junior members you know, kind of young people from the company might want to have a more kind of colloquial kind of tone of voice and maybe more senior partners and executives would want a more formal one'. From my research, this personalisation of voice responses would require profiling to understand the human who is speaking but also to define the profile of the synthesised voice output. This move to create greater personalisation would require far more in-depth and ubiquitous vocal profiling. It was identified as market-driven: the same participant said that, 'clients now want to personalise experience'.

Participants noted that they 'always try to push [clients] to use a gender-neutral name' for their voice assistant, suggesting that there is growing concern about addressing gender bias in the field of AI. However, the remarks call attention to the significant blind spot obscured by dominant understandings of voice as an indicator of identity and personhood, which, when escalated, incorporates broader issues around profiling far beyond gender. I noted that this understanding is probably heavily influenced by the participants' backgrounds in linguistics and data analytics. The comments signified how my proposed Speculative Voicing Framework could be used for vocal imaginaries with an ecological and social sensibility, but, equally could be capitalised on for product development. Ultimately, my project motivations are fairly experimental, whereas the motivations of IBM employees are driven by practical application.

When asked to give feedback on the overall workshop experience via a short Google Forms questionnaire (Appendix D), participants were seen to take away some key features of the concept of Speculative Voicing that I had presented to them. One participant commented that the workshop had made them think differently about voice, with 'the idea that a single person may have multiple voices and the usage of different voices for different contexts'. They added that, the workshop had made them consider voice with 'more importance on the actual sound and it's characteristics as opposed to the words uttered'. Another participant noted, 'I think I've not really considered the "noises" to be part of voice, or the entire audio experience'.

While the findings presented above show a degree of success in the micro-environment of this PhD, at the macro level they have also highlighted the need for more interdisciplinary work to divert concentrated attention away from normative and over-simplistic profiling practices when defining voices. As intended with this research inquiry, speculative design could play a crucial role in presenting more nuanced notions of voice, and the methodology described and presented in this thesis could be a means to enact this. I am working to initiate more curiosity and involvement from the creative community to engage with voice as material. This has been activated through my engagement with artists, designers and scholars via guest lectures and symposia (See: Appendix E). I have also set up the Instagram account @Speculative_Voicing (Abbas-Nazari, n.d.), which aims to catalogue projects 'exploring the potential of the sounded voice as a material',¹⁶¹ to catalyse further discussion around the sound and sounding of voices beyond profiling.

Conclusion

This evaluative activity shows promising results for Speculative Voicing, as a concept, to explore the multiplicity of voice in relation to being and identity in a

¹⁶¹ Projects, many of which are referenced in this thesis.

conversational AI context, to reveal and resist voice profiling. Participants conjured their own auditory insights and vocal imaginaries from the audio I played to them, evidenced by their descriptions of what they heard. The analysis of the evaluative findings (Figure 45) provides evidence to present to AI industry professionals, such as those at IBM, and advocacy groups to initiate discussion around recommendations for revising methods of working with voices in conversational AI that do not rely on harmful profiling practices.

The context of the practice as socially and ecologically entangled failed to resonate fully with the participants. It indicates that for the projects to reach their greatest potential, they cannot be presented purely as audio clips but must be accompanied by their intended supporting materials and designed interactions that provide narrative and storytelling elements to the projects. Nevertheless, presenting the audio clips in isolation was essential in order not to influence participants in the evaluative activity.

It is important to note the small sample size in evaluating this work. However, the participants also represent the key players working in the field that this research addresses. Continuing with the line of enquiry initiated in this PhD research, on a larger scale, I am confident that this work can make significant ‘understanding and awareness, attitudinal and cultural impacts’ (Reed, 2018) to reveal and resist vocal profiling practices in conversational AI systems.

When questioned, participants returned to their dominant understanding of voice and voicing. At the time of writing a reciprocal and corresponding relationship of understanding between human and synthesised modes of vocal sounding, exists.¹⁶² With increased creative and speculative exploration of the sound and sounding of voices, as proposed, this could have a significant effect in disturbing

¹⁶² See Chapter 1 for further discussion.

dominant understandings of the ability of voice to be profiled. In the concluding chapter of this thesis, I present a workshop that prototypes a response to this key finding and the prevailing issue exposed by this evaluation.

Chapter 8: Conclusion

Introduction

This practice-led research aimed to develop discussion about and challenge how AI and the AI industry comprehended the sound and sounding of voices. Taking a position from music and speculative design practice, I contested ways of working in conversational AI systems to shape notions of (vocal) identity. I followed a practice-led methodology of sonic speculative design, which emerged and was developed during this research. This chapter describes the research outputs and their intended multiple audiences, commencing with discussion of a final workshop that marks the culmination and conclusion of this PhD research. In addition, I discuss the findings of the research questions, original contributions to knowledge and proposed future work that has emerged from this investigation.

Outputs

Speculative Voicing Workshop (for AI Industry Professionals)

For human voices, a main obstacle to revealing and resisting vocal profiling in conversational AI systems is the dominance of frameworks underpinned by concepts from the field of linguistics and phonetics. It maintains that vocal sound alone can substantiate the speaker's mental state, physiology and anatomy (Müller, 2007). This was also demonstrated in the IBM Workshop: participants stated that when designing voice-user interfaces and voices of conversational AI systems, they ask clients to think about these voices as people.¹⁶³

I addressed this fundamental barrier to appreciating vocal materiality and enabling vocal profiling through a 1.5-hour 'Speculative Voicing Workshop' (Figure

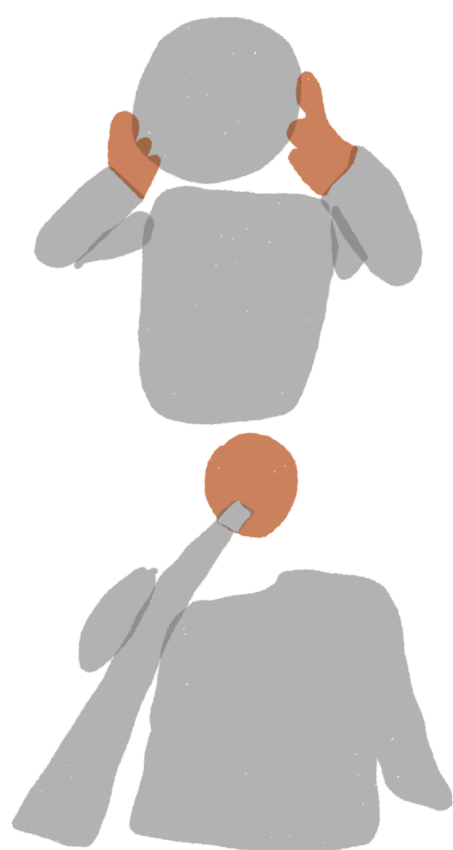
¹⁶³ See previous chapter.

46) with attendees of the 2023 Articulating Data symposium in Edinburgh.¹⁶⁴ It was developed using the lessons I had learnt from the ‘Speculative Listening’ workshops, with the addition of the understanding I had gained from my two case study projects and their evaluation by IBM employees.

This workshop was devised for those working or intending to work with voices in conversational AI systems, such as those I met during the IBM workshop. It is intended to be deployed to unlearn the dominant conditions of vocal profiling, using the Speculative Voicing methodology. The workshop provides the opportunity to explore an experience of vocal materiality and polyphony first-hand, which underpins the Speculative Voicing methodology. It could be utilised as a precursor to generate methods that do not rely on vocal profiling for designing and defining synthesised and human voices, respectively, supported by analysis of the evaluative findings of the Speculative Voicing Framework (Figure 45).¹⁶⁵

¹⁶⁴ The Articulating Data symposium brought together people, practitioners and researchers interested in ‘vocalisation, machine listening, and the (in)security of language in a digital age’ (Articulating Data, 2023). I also gave a short talk about my PhD research on this occasion.

¹⁶⁵ See Evaluation chapter.



DIY
voice modification
devices

mixed muddy smooth
scooping wooden narrow
stressed wobbly yawny
strangled husky forced

Figure 46: Speculative Voicing Workshop. Elspeth Murray

In the workshop we made DIY voice modification devices, as enacted in the ‘Polyphonic Embodiment(s)’ practice project. The workshop was documented through live drawing by Elspeth Murray and a short video by the symposium’s organisers ([Item 30](#)). To guide the participants, they were invited to collect a voice quality descriptor term from a ‘lucky dip’ to then think about how they could modify their voice to sound like the descriptive term received (Figure 46). The voice quality descriptor terms were obtained from *Profiling Humans from their Voices* (Singh, 2019, pp. 242-251) (Figure 47), as were those used in the ‘Polyphonic Embodiment(s)’ project. As with the other workshops I have initiated as part of the PhD research, participants were provided with simple, readily available materials to create their designs. During the workshop activity I encouraged participants to focus on vocal materiality, why and how our bodies make the vocal sound they do, thinking about whether they have multiple voices and when and how they enact

them. They were also asked to consider the subjectivity contained within the descriptive term they had collected.

<i>1 Airy</i>	<i>2 Animated</i>	<i>3 Aphonic</i>	<i>4 Babyish</i>	<i>5 Back</i>
<i>6 Balanced</i>	<i>7 Beautiful</i>	<i>8 Belted</i>	<i>9 Biphonic</i>	<i>10 Biting</i>
<i>11 Bleating</i>	<i>12 Brassy</i>	<i>13 Breathly</i>	<i>14 Bright</i>	<i>15 Brilliant</i>
<i>16 Broad</i>	<i>17 Buzzy</i>	<i>18 Chesty</i>	<i>19 Clear</i>	<i>20 Cloudy</i>
<i>21 Coarse</i>	<i>22 Cold</i>	<i>23 Constricted</i>	<i>24 Covered</i>	<i>25 Crackly</i>
<i>26 Creaky</i>	<i>27 Dark</i>	<i>28 Denasal</i>	<i>29 Diplophonic</i>	<i>30 Dulcet</i>
<i>31 Dull</i>	<i>32 Easy</i>	<i>33 Efficient</i>	<i>34 Falsetto</i>	<i>35 Firm</i>
<i>36 Flageolet</i>	<i>37 Flat</i>	<i>38 Flutelike</i>	<i>39 Focussed</i>	<i>40 Forced</i>
<i>41 Front</i>	<i>42 Full</i>	<i>43 Glottal</i>	<i>44 Grating</i>	<i>45 Gravelly</i>
<i>46 Gruff</i>	<i>47 Guttural</i>	<i>48 Harsh</i>	<i>49 Head</i>	<i>50 Heavy</i>
<i>51 High</i>	<i>52 Hoarse</i>	<i>53 Honky</i>	<i>54 Hushed</i>	<i>55 Husky</i>
<i>56 Icy</i>	<i>57 Intense</i>	<i>58 Labored</i>	<i>59 Lilting</i>	<i>60 Limpid</i>
<i>61 Loud</i>	<i>62 Low</i>	<i>63 Luxuriant</i>	<i>64 Mellow</i>	<i>65 Melodious</i>
<i>66 Metallic</i>	<i>67 Middle</i>	<i>68 Mixed</i>	<i>69 Modal</i>	<i>70 Modulated</i>
<i>71 Monotonous</i>	<i>72 Mouth</i>	<i>73 Muddy</i>	<i>74 Muffled</i>	<i>75 Musical</i>
<i>76 Naked</i>	<i>77 Narrow</i>	<i>78 Nasal</i>	<i>79 Noisy</i>	<i>80 Normal</i>
<i>81 Open</i>	<i>82 Opera</i>	<i>83 Pale</i>	<i>84 Pleasant</i>	<i>85 Pressed</i>
<i>86 Pulsed</i>	<i>87 Raspy</i>	<i>88 Relaxed</i>	<i>89 Resonant</i>	<i>90 Retracted</i>
<i>91 Rich</i>	<i>92 Ringing</i>	<i>93 Rough</i>	<i>94 Rounded</i>	<i>95 Scooping</i>
<i>96 Shaky</i>	<i>97 Shrill</i>	<i>98 Slack</i>	<i>99 Sliding</i>	<i>100 Smooth</i>
<i>101 Soft</i>	<i>102 Steady</i>	<i>103 Stiff</i>	<i>104 Strained</i>	<i>105 Strangled</i>
<i>106 Stressed</i>	<i>107 Strident</i>	<i>108 Stroh bass</i>	<i>109 Strong</i>	<i>110 Sugary</i>
<i>111 Sultry</i>	<i>112 Supported</i>	<i>113 Sweet</i>	<i>114 Tense</i>	<i>115 Thick</i>
<i>116 Thin</i>	<i>117 Throaty</i>	<i>118 Tight</i>	<i>119 Tinny</i>	<i>120 Tremorous</i>
<i>121 Twangy</i>	<i>122 Ugly</i>	<i>123 Unforced</i>	<i>124 Unpleasant</i>	<i>125 Ventricular</i>
<i>126 Vibrant</i>	<i>127 Warm</i>	<i>128 Weak</i>	<i>129 Whining</i>	<i>130 Whistly</i>
<i>131 Wobbly</i>	<i>132 Wooden</i>	<i>133 Wooley</i>	<i>134 Yawny</i>	

Figure 47: Voice Quality descriptor terms. Adapted from (Singh, 2019, p. 242).

The vocal quality descriptors that were designed into DIY voice modification devices included ‘strangled’ (Figures 48 & 49), ‘modal’ (Figure 50), ‘broad’ (Figure 51), ‘low’ (Figure 52), and ‘intense’ (Figure 53). Participants were invited to listen to a

playlist of recordings by female experimental vocalists, including Cathy Berberian, Meara O'Reilly, Holly Herndon and Laurie Anderson, all referred to in this thesis. As with the other workshops I conducted during this PhD, the playlist was to encourage participants to evolve ideas catalysed from a position of sound and sounding, but in this case specific to voice. Participants were informed that if they didn't feel comfortable putting materials on their body or face, they could alternatively make something and imagine how it would augment or transform their body and vocal sound. They were also instructed that if the ideal material to make their idea was not among that provided, this could be illustrated using an alternative material that was available.



Figure 48: Speculative Voicing Workshop: strangled.
Elspeth Murray.



Figure 49: Speculative Voicing Workshop: strangled.
Elspeth Murray.



Figure 50: Speculative Voicing Workshop: modal. Elspeth Murray.



Figure 51: Speculative Voicing Workshop: broad. Elspeth Murray.

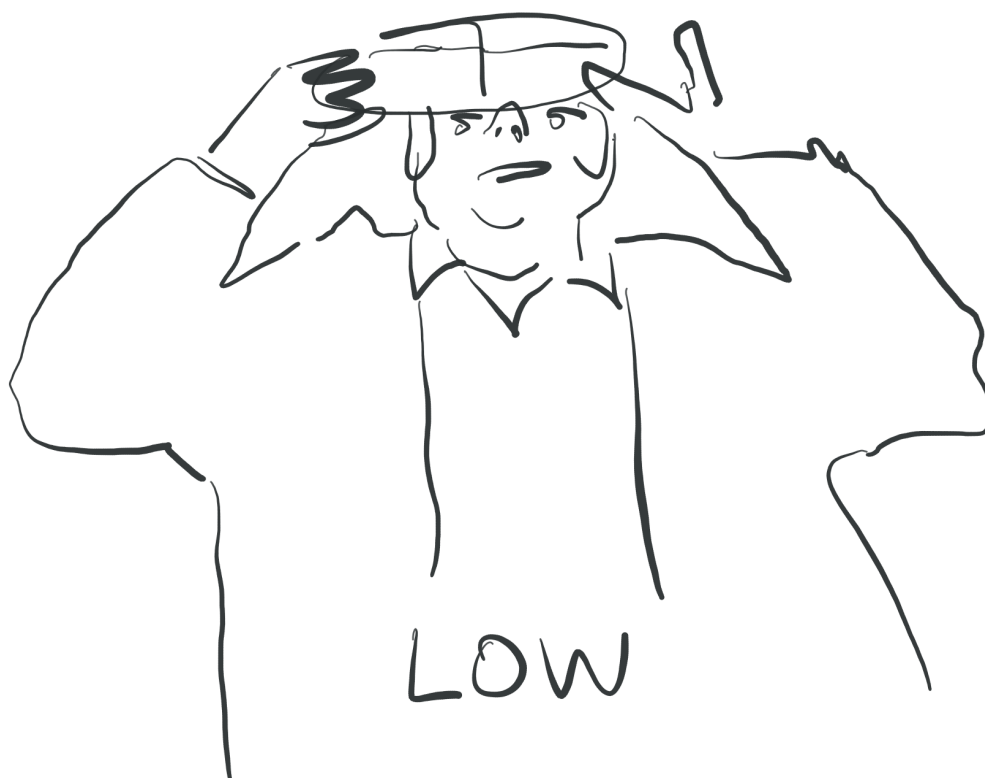


Figure 52: Speculative Voicing Workshop: low. Elspeth Murray.



Figure 53: Speculative Voicing Workshop: intense. Elspeth Murray

Speculative Voicing Framework (for Speculative Designers and Advocacy Groups)

This PhD research proposes that the field of AI could benefit from greater interdisciplinary work and incorporation of ideas from research and practice by those working speculatively with design, technology and/or sound, to generate alternatives to vocal profiling and to dismantle AI's dominance in defining understandings of voices. The Speculative Voicing Framework, developed in Chapter 4, and tested and exemplified in Chapters 5 & 6, provides written guidance on applying the sonic centric methodology to voices in conversational AI systems, in ways which reveal and resist vocal profiling.

Through evaluation and analysis of the Speculative Voicing methodology, in Chapter 7, thematic features of voice profiling in comparison to speculative voicing was generated. This emergent data provides useful information for speculative designers and advocacy groups working to renew understandings of voices in AI and more responsible working methods.

Interactive Tools (for Speculative Designers and Non-Specialists)

The research has produced two interactive tools to support the application of the Speculative Voicing Framework and further the production of speculative voices and vocal imaginaries. The first of the two tools is the [wav2face Google Colab \(Item 19\)](#), a voice-to-face AI recognition tool created as part of 'Polyphonic Embodiment(s)', which is made available as an application for other researchers and practitioners to experiment with. This tool reveals implicit assumptions in voice profiling to produce facial images.

The other tool was developed and evolved from the 'Acoustic Ecology of an AI System' project and follows on from work initiated by female experimental vocalists, as this thesis explored. It provides an interactive resource for prototyping new voices

using one's own voice. This tool is created with the software Max/MSP (Cycling '74, n.d.) to create a 'patch' and allows individuals to use their voice as an input and modify it in real time using additional plug-ins.¹⁶⁶ I created the initial prototype (Figure 54 / [Item 31](#)) and then I sought the expertise of Andy Sheen to create the final version (Preview - [Item 32](#) / Patch - [Item 33](#)).

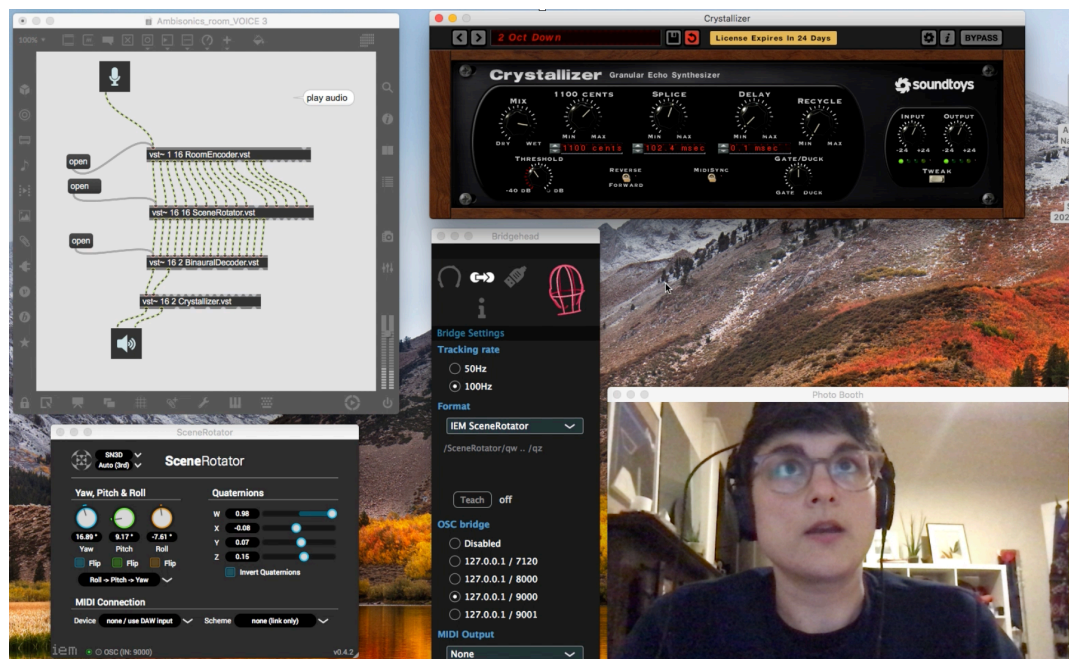


Figure 54: Screenshot of video documenting prototype version of Max/MSP patch

For synthesised voices, the requirement for 'intelligibility and naturalness' (Sutton et al. 2019) in conversational AI remains a strong barrier to incorporating more narrative and storytelling sonic elements into the sound and sounding of these voices. As with 'Acoustic Ecology of an AI System', the Max/MSP¹⁶⁷ tool takes a vococentric and voice-led approach to manipulate the voice's sonic material and

¹⁶⁶ The plug-ins I used were Crystallizer (Soundtoys, n.d.) and Crowd Chamber (QuikQuak, n.d.), as they both provide much opportunity to modify and experiment with the voice and consider its materiality in relation to associated factors. Crowd Chamber appears in the original project, 'Acoustic Ecology of an AI System'. However Rayspace (QuikQuak, n.d.) was discontinued when producing the Max/MSP patch, so Crystallizer was used as an alternative. The open nature of a Max/MSP patch also means that these plug-ins could be swapped according to someone else's investigation or enquiry into voice modification.

¹⁶⁷ In my other works and outputs, I have used free, open-source software, where possible. Max/MSP and the audio plug-ins I utilise here are not free, but they are easily accessible, and Max/MSP is very open-ended in the possibilities it allows. A future aim would be to find a way to make the 'Voicing Beyond the Vacuum' tool more accessible, perhaps by using the Pure Data software, which is an open-source visual programming language for multimedia.

materiality, unrestricted by current AI processes, reimagining voices in conversational AI systems.¹⁶⁸ Here, a merging of human and synthesised voice emerges for others to work with and contemplate the sound and sounding of voices in conversational AI systems. Called, ‘Voicing Beyond the Vacuum’, this open-ended, playful tool encourages others to explore their own experimental vocal polyphonic potential to resist vocal profiling (Figure 55). It enables the consideration and contemplation of alternative modes of voicing for storytelling and creating vocal imaginaries, as is intended with the Speculative Voicing methodology defined through this PhD research.

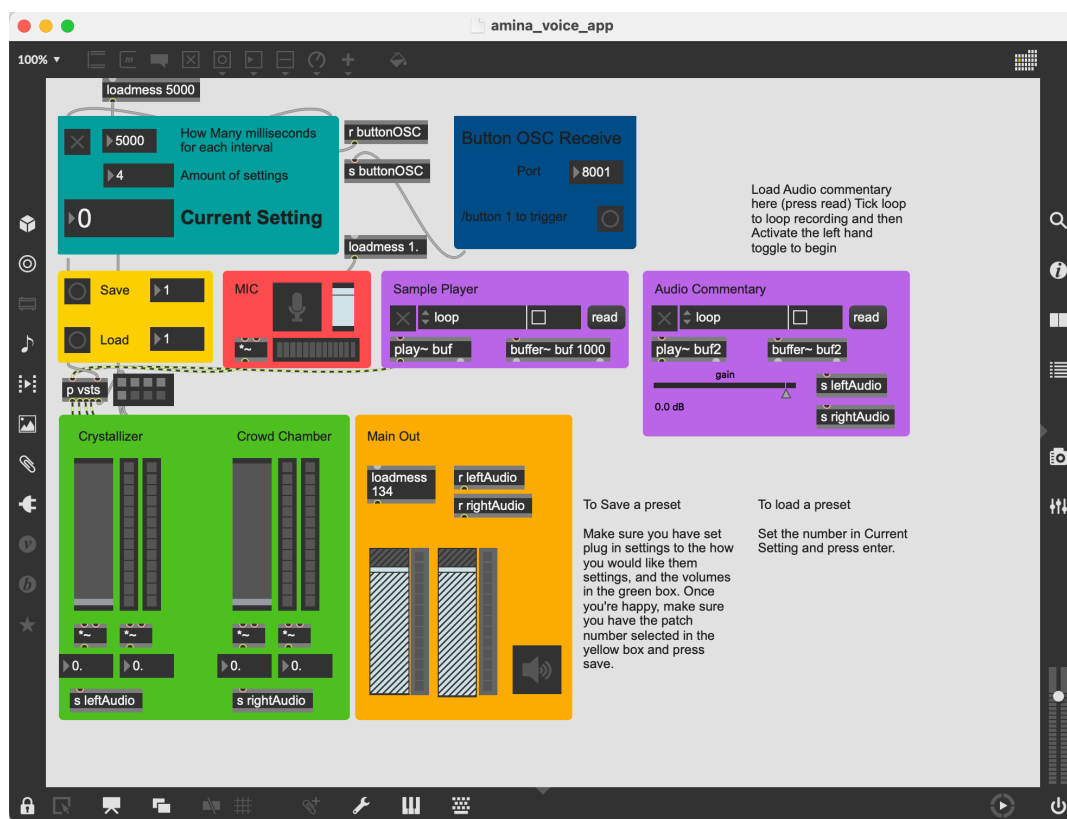


Figure 55: Screenshot of final Max/MSP patch. Amina Abbas-Nazari & Andy Sheen

These two available tools allow people to carry out a feedback process of sounding speculative voices to reveal and resist vocal profiling. As with the Speculative Voicing Framework, they provide easily approachable and accessible

¹⁶⁸ There is the option to include head-tracking using a small piece of hardware and Bridgehead software made by Supperware (n.d) to create immersive experiences using the ‘patch’ created.

resources to enact the sonic speculative design methodology. They might be utilised by speculative designers, community groups or creatives to continue to explore vocal potential and refute AI's profiling practices. Collectively, the tools provide robust resources for those working speculatively to apply the Speculative Voicing methodology independently.

Documentation (for the Public)

Documentation of the case study and workshop series practice projects are detailed in a research output for public dissemination on the Speculative Voicing website (Abbas-Nazari, 2022) (Figure 56). In addition, the Speculative Voicing Instagram page (Abbas-Nazari, n.d) documents and catalogues wider contextual research from the project that highlights other practitioners' use of voice as a sonic material within different creative realms of work, such as music, technology, art and architecture. These resources seek to educate and inform a broad audience about vocal expression beyond voice profiling practices. These outputs support the dissemination of this research, future research and intend to increase its impact.

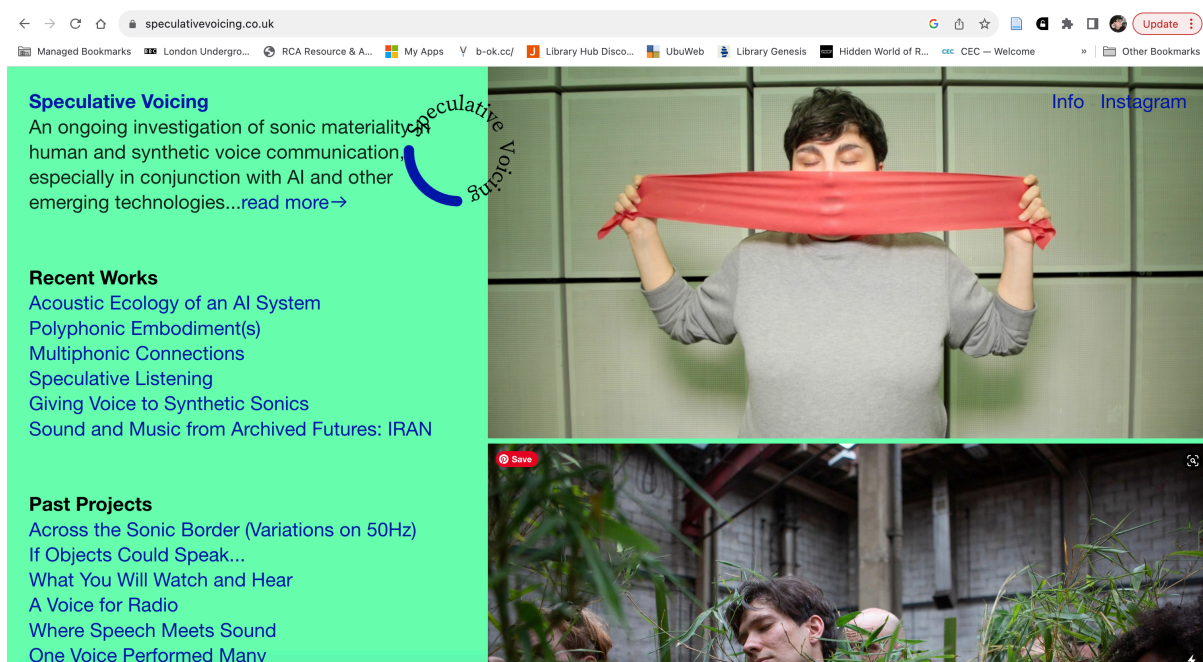


Figure 56: Screenshot of Speculative Voicing Website.

Findings

In this thesis, I posed three questions. In answer to the questions, I summerise:

1. How can thinking with and through sound develop a sonic speculative design methodology?

In Chapter 3, I investigated the potential of sonic thinking combined with speculative design through six workshops with young people or those working with young people. The use of sonic thinking to drive and catalyse speculative design practice demonstrated that this produced ideas and concepts of co-creation. A sonic speculative design methodology developed, which built upon and evolved existing theory, providing an intersectional new materialist position (Chapter 4). The methodology was distilled into a Speculative Voicing Framework as a research output, and to apply and test the methodology in two case study practice projects (Chapters 5 & 6). Applying sonic thinking to speculative design advanced a methodology which, when evaluated in the 'IBM workshop', provided findings to inform more conscientious conversational AI development (Chapter 7). A final 'Speculative Voicing Workshop' (Chapter 8) was prototyped to further procure this for future work and impact.

2. What does applying this methodology reveal about vocal profiling, in the AI era?

Having established the negative impact of voice profiling practices in conversational AI systems (Introduction & Chapter 1), I sort to critique these methods from the position of sounding, as opposed to listening, utilising the Speculative Voicing methodology. Applying the methodology allowed for a comparison of the way voices are currently conditioned to reveal vocal profiling practices in conversational AI systems and the themes which underpin this conditioning (Chapter 7) The friction between the two modes of working with and understanding voice generated critique and vocal imaginaries to illustrate

conversational AI systems in more holistic ways ('Acoustic Ecology of an AI System'). The methodology challenged assumptions of sounded voices in conversational AI concerning facial identity ('Polyphonic Embodiments'). The wav2face Google Colab tool provides a research output to further reveal and explore vocal profiling.

The IBM Workshop (Chapter 7) showed that deeply rooted relations between voice and profiling remain. As a result, the 'Speculative Voicing Workshop' was devised to respond to this finding (Chapter 8).

The theoretical research revealed that vocal profiling in conversational AI systems is enacted through modes of reasoning that pertain to the voice as though it exists in a vacuum, and that this, in turn, actively neglects the materiality and polyphonic potential of voice.

3. How does applying this methodology resist vocal profiling in the AI era?

The Introduction and Chapter 1 identified that voice profiling aims to correlate vocal sonic features to determine wide-ranging personal attributes of humans or imagined human personas, which relies on normative expectations and oversimplistic notions. Speculative Voicing was shown to successfully resist this. In Chapters 5 and 6, the PhD's practice component demonstrated that a polyphonic understanding of the materiality of voice via the perspective of 'sounding' renders practices of profiling untenable.

The effectiveness of resisting vocal profiling was evidenced through evaluation of the work during the IBM Workshop (Chapter 7), in which participants no longer tried to describe and conceptualise one person in relation to vocal sounding but instead produced vocal imaginaries. The 'Voicing Beyond the Vacuum' tool and the

‘Speculative Voicing Workshop’ provide research outputs to further support resistance to vocal profiling through the continued creation and exploration of speculative voices and vocal imaginaries by others.

Limitations and Considerations

It is important to note that this PhD research has been conducted from an intersectional position, motivated by social and ecological standpoints. In the wider real-world context, this research could potentially be used and abused in other ways, outside and beyond the author’s control. All forms of knowledge and technology can be both tools and weapons, and often simultaneously, especially when in the hands of humans. For example, fire can be used to cook nutritious food, but equally it can be used to destroy the infrastructure that sustains humans.¹⁶⁹ I sincerely hope that the people who take an interest in this research exploit it for the purposes for which it was intended.

Since this research was initiated in September 2018, interest in AI has been continually growing. In 2023, OpenAI released ChatGTP (2023), an AI-powered language model that generates human-like text responses and the year before, DALL-E 2 (2022), an AI system that can create realistic images and art using simple text descriptions. It is becoming clear that AI is permeating new areas of the socio-cultural fabric daily. This is evidenced by the recent strikes by the Screen Actors Guild-American Federation of Television and Radio Artists (Sag-Aftra) and the Writers Guild of America (WGA) against the use of deepfake AI technology in the film industry (Beckett & Paul, 2023).

The momentum of AI expansion and development is not helped by high-profile but unrealistic, irresponsible representations of AI. For example, a former Google employee claimed that an AI chatbot had become sentient (See: Lemoine, 2022).

¹⁶⁹ See also: (Pohflepp, n.d.)

While ethicists can now be found working at all major technology companies, they risk potentially being fired for speaking out about harmful practices in AI. This was the case with Timnit Gebru, who was expelled from Google for highlighting the risks of large language models (See: Hao, 2020). With the persistence of AI innovation, and despite the challenges, it remains essential that AI frameworks, practices and uses are continually called into question. With this PhD research, I see the potential to stimulate creative interest in the voice as material, to explore concepts of being and identity, and engage with emerging issues in AI. In turn, this can provide new methods for including a broader range of people, theorists and practitioners who perhaps might not have engaged with AI before, in the discussions of AI ethics.

Voice could be considered the ultimate, original form of human communication. However, its sound and sounding are not yet fully appreciated in design research and practice as a material form that has the polyphonic potential to be designed. This PhD research creates opportunities to work with voice while also addressing some longstanding issues in speculative design and emerging issues in AI by applying sonic thinking.

Original Contributions to Knowledge

This PhD research provides a sonic speculative design methodology as its core original contribution to knowledge. It is accompanied by tools and resources to enact the methodology. The methodology advances speculative design, providing an intersectional, new materialist position via a novel relationship with sonic thinking, demonstrated through the theory and practice in this PhD.

In the context of this PhD, Speculative Voicing was deployed to reveal and resist voice profiling, building on the critique of voice profiling in conversational AI as normative and marginalising from the perspective of sounding.

Supplementary contributions to knowledge provide a body of practice-based work that demonstrates the voice as a material for speculative design practice, and case studies explored how voice could be utilised for depicting vocal imaginaries, via the adopted methodology.

Future Work

Advocacy Work

This research was about developing a sonic speculative design methodology and how this could reveal and resist AI voice profiling. As a result of the research, emergent knowledge has been generated which could be useful for advocating against these harmful practices and for empowering communities through bottom-up initiatives. Initially, I hope to compile this PhD research's major findings and key outputs into an easily accessible format such as a downloadable or print-on-demand book. My intention is to ensure that the resources and knowledge that reveal and resist vocal profiling that have been generated by this research can effectively reach communities which are currently negatively affected, to advocate for more conscientious, responsible AI development. I am also hoping to return to IBM to conduct a Speculative Voicing workshop, communicate the outcomes of the research and provide analysis of the evaluative process they participated in. Potentially we could work together to create recommendations for revising methods of working with voices in conversational AI that do not rely on profiling practices. I will also continue to offer participatory speculative voicing workshops to the increase dissemination and impact of this research.

Theoretical Work

I plan to continue developing the Speculative Voicing methodology by exploring the synergy of sound and design practice in more depth. I want to explore how the methodology can be expanded and applied beyond the context of voices in conversational AI systems. I would also like to further establish an intersectional

position for speculative design. Gathering qualitative data during future Speculative Voicing and related workshops could be beneficial for this. When conducting most of my workshops, for example the 'Speculative Listening' series, I struggled to identify the questions to ask participants which would yield useful data for this study. Instead, my research was unequivocally led by practice, observation, and 'reflection on/in action' (Schön, 2016). If I had questioned participants, this would have probably resulted in a different outcome for this project. Now that I have established original contributions to knowledge, the future dissemination and publication of the research could be supported by additional participant feedback.

The areas of this thesis that centred on discussing the metaphor of a vacuum to draw attention to the materiality of voice, when extrapolated, ultimately leads to the acknowledgement that artificial intelligence is not alive. It does not and cannot breathe. While too lengthy and outside the remit of this investigation to explore in depth, the metaphor could be extended to provide a useful context for conversations around AI, cognition and consciousness.

Practice

So far, the case study projects have been presented online and through symposiums and guest lectures. They were also presented in the IBM Workshop setting, which allowed me to gather valuable feedback and evaluative information. In the future, I hope to display the case study projects to public audiences during festivals and exhibitions in order to engage a broader audience with the themes and findings of this work. These occasions could also provide further opportunities to gather evaluative data, similarly to the IBM Workshop.

New Challenges

New challenges in AI are emerging at a rapid rate. I hope this thesis elevates the importance of the sound and sounding of voices and vocal profiling to gain levels of

traction, interest and critique that compare to those of its visual counterpart, facial recognition. The need for this is urgent, with the currently increasingly blurred line emerging between human and synthesised voices via voice cloning apps and deepfake processes (See: Edwards, 2023). Again, this particular issue is too large to be fully incorporated into the scope of this PhD investigation, but it may be the next challenge to confront.

Appendices

Appendix A:

'Inaudible Audio' Track Listing

(Text in italics - narration by synthesised AI clone of my voice)

This is the sound of a baby in its mothers womb

HouseOFMeis (2016) FINDING BABIES HEARTBEAT AT 10 WEEKS! HOME DOPPLER!
<https://www.youtube.com/watch?v=f80rbeRjJkc>.

This is the sound of bats

justsoundfx (2013) Bats Sound. <https://www.youtube.com/watch?v=ppLsu5Z2Np0>.

This is the sound of corn growing

UNL CropWatch (2016) Listen and Watch Corn Grow. <https://www.youtube.com/watch?v=76xEkEXI2a4>

This is sound from the Sun

NASA (2018) NASA | Sun Sonification (raw audio). https://www.youtube.com/watch?v=-I-zdmg_Dno.

This is the sound of your stomach digesting food

stomach and intestines sound (2018) 【belly noises】 Active stomach and small intestine after meals. <https://www.youtube.com/watch?v=-IL3ky-Kcj8>.

This is the sound of an aircraft travelling at the speed of sound

MW Hub (2017) SONIC BOOMS & JETS | Best Compilation. <https://www.youtube.com/watch?v=jmhU7SEo4gg>.

This is the sound of a vibrating Antarctic iceberg

AGU (2018) Ghostly sounds of a vibrating Antarctic Ice Shelf. <https://www.youtube.com/watch?v=w56RxaX9THY>.

This is the sound of your heart beating

justsoundfx (2013) Heartbeat sound. https://www.youtube.com/watch?v=gJpT_wHZeF8.

This is the sound of a low frequency underwater sonar and whale calling

Deep Blue Sea (2015) Modern Sonar Sounds and other Sounds of the Sea. <https://www.youtube.com/watch?v=fXfNfvnDyQI>.

This is the sound of Saturn's radio wave emissions

Space Audio (2014) Cassini RPWS: Bizarre Features of Saturn's Radio Emissions. <https://www.youtube.com/watch?v=66afDxU1Ilg>.

This is the sound of electricity

vina54 (2017) 50 Hz vs 60 Hz vs 400 hz A.C. Hum Sound Comparision. <https://www.youtube.com/watch?v=pMtn-loUrg8>.

This is the sound of a mixed frequency underwater sonar

Deep Blue Sea (2015) Modern Sonar Sounds and other Sounds of the Sea. <https://www.youtube.com/watch?v=fXfNfvnDyQI>.

This is the sound of radio-waves in the earths' atmosphere

NASA (2014) Chorus Radio Waves within Earth's Atmosphere. <https://soundcloud.com/nasa/chorus-radio-waves-within-earths-atmosphere>.

This is the sound of rhubarb growing

Anon (2011) A mass of popping rhubarb - forced rhubarb growing at DWS Farm. <https://soundcloud.com/rhubarb-rhubarb-rhubarb/a-mass-of-popping-rhubarb>.

This is the sound of long-finned pilot whales

Deep Blue Sea (2015) Modern Sonar Sounds and other Sounds of the Sea. <https://www.youtube.com/watch?v=fXfNfvnDyQI>.

This is the sound of deep cosmic background radiation

New Scientist (2014) Recording captures hiss of Big Bang radiation. <https://www.youtube.com/watch?v=gJJmFnMea1Q>.

This is the sound of lightening on Jupiter

Ordo Science & Tech (2018) The Sound of Universe: Voyager - Lightning on Jupiter. <https://www.youtube.com/watch?v=K5YZX3xnKsc>.

This is the sound of ice skating on thin ice

National Geographic (2018) Hear the Otherworldly Sounds of Skating on Thin Ice. <https://www.youtube.com/watch?v=v3O9vNi-dkA>.

Appendix B:

Multiphonic Connections Automated Telephone Script

(Call Received)

Call Introduction

Thank you for calling Multiphonic Connections

We make the inaudible audible.

Our collection of sonic experiences include those which are only audible to you through the use of technology. Choose an option you would like to hear.

Press zero for your initial listening calibration exercise

Press 1 for sounds from inside the human body (Audio Playlist)

2 for far, far away, cosmic sounds (Audio Playlist)

3 for sounds from deep below you (Audio Playlist)

4 for sounds made by animals (Audio Playlist)

5 for sounds that manifest very slowly (Audio Playlist)

6 for sounds that happen extremely fast (Audio Playlist)

7 for very quiet sounds (Audio Playlist)

Press star at anytime to return to this menu

And 9 at anytime to leave a voicemail message

(Zero Pressed)

Initial Listening Calibration Exercise

Are you listening? Are you listening to my voice? My voice is synthetically created using technology.

During this call my voice will connect your ear to different sonic experiences. We will first explore listening experiences enabled by your human anatomy. Then we'll experience a range of sounds that you, as a human, can only hear through the use of technology. Finally, Multiphonic Connections will invite you to request speculative or imaginary sounds to be added to our ever growing collection of sonic experiences.

Since joining this call your auditory attention has been focused on the sound of my voice but there are many other sounds unfolding around you that you can also attune to.

Broaden your listening perspective to incorporate other sounds around you: perhaps you can hear traffic buzzing or trees blowing in the wind?

What is the loudest sound you can hear at this moment?

Is it high or low pitched?

What is the most prominent or pervasive sound you hear?

Find the most distant sound you can hear? How far away do you think it is?

Pinpoint the quietest sound you can find in your listening experience?

What is the sound you hear closest to you?

As you breathe, listen to the sound of your breathing. Your breathing connects your inner and outer worlds of perception.

As you breathe in consider the sounds being made internally by your body. If your surrounding environment is very quiet you may be able to hear your heart beating or your stomach working to digest food.

The sounds humans are able to hear typically fall into the range of 20 to 20,000 Hertz, although there is considerable variation between human individuals.

There are many sounds humans are unable to hear with your ears alone because they might:

Happen too fast.

Manifest too slowly.

Are too high frequency,

or too low frequency.

Some sounds are just too far away from your proximity,

too quiet

or too far below the ground for you to perceive.

Then there are sounds that happen internally, inside your body

From the stethoscope to satellites - Technology allows humans to hear things not normally audible to you. Electronic communication has made us aware that once silent domains are in fact spheres of sound and noise. We have extended the range of human hearing as never before.

Press star at any time to return to the main menu.

And 9 at anytime to leave a voicemail message

(9 Pressed)

Voicemail Message Instructions

Please note your voicemail is being recorded and will contribute to the Empathy Loading programme and our research purposes.






Multiphonic Connections is making the in-audible audible. Multiphonic Connections creates sonic experiences to enable you the listening abilities to hear sounds not normally audible to humans.

But, how do you want to be heard by technology? How can technology allow you to be heard and listen in different ways? How does what you hear shape the way you understand the world and other beings? There are still many, many sounds that have not yet reached the human ear.

We invite you to reflect on these questions and request speculative or imaginary sounds to be added to our collection of sonic experiences. Leave a voicemail message with your thoughts. We look forward to hearing your suggestions. Please leave your message after the tone and hang up when you have finished.

Appendix C:

Mural Board from IBM Workshop ([Item 29](#))

PARTICIPANT	TEST	1	2	3	4	5	6	7	8		
M.	 a conference with a body set up sound system. More. Does the volume kick in a little later? Content opens. feels like it would be intrusive and factual (from the tone)	echo, witchy, mid-heavy (feels natural), volume falls overtime.	choppy, mechanical, speaking through a fan, reverb, square with rounded corners, a sick robot.	atmospheric with a defined speaker, starting of sounds the speaker and the atmosphere are separated, feels like a complete low level, almost ambient, sometimes doctor and some atmosphere. Dystopian?	strangely rhythmic, echo feels at a slightly reducing tempo with a slightly increasing pitch. Like the opposite of skipping a stone	a conference with a body set up sound system. More. Does the volume kick in a little later? Content opens. feels like it would be intrusive and factual (from the tone)	intonations feel distinct, very short resonance, ASMR-ish subtle wellness/click	crispy reverb, very delayed single echowith loads of distortion, slowed down, lowered pitch,	genuinely thought someone was asking a question, sounds from the front of the mouth, constant female voice		
O.	 faint creaking of one of the other doors in the house. creak intermittently, twice each time sounds a bit like going up and down like this	female voice, unclear, echo-y (with strong echo) disconcerting, feel a bit uncomfortable hearing it	distorted sound, can't identify speaker, ends in fading hum	speaker (female) voice against otherwise fairly background noise which is being being in an airport with airport announcements although what she is saying almost sound like an announcement. it's more like news report	echo-y on short segments of the speech, so it seems less natural / more techy	sounds like a male voice, cannot understand/make out language, fairly clear despite some added static/distortion, like the voice/intonation	female voice with echobut off, like on calls with annoying connection, sounds like normal speech	background noise slightly quieter & louder, reminds me of threatening scenes in thrillers, female voice speaking calmly	sounds like the speaker has a stuffy nose, slightly hoarse voice, can't understand what they're saying, female voice		
S.	 very faint but with the same rhythm sounds like tick, raising a small tick (my metal cap of my sweater under the washing machine)	far away, loong, tired, last part feels like they are outside	scary, sinister sound, 	sounds like a rocket ship launch announcement, someone making an important announcement to many people but interrupted by loud connection due to too high pitch (British accent?)	choppy, double echo feels like 3 different people speaking, the person seemed sad 	sounds like a guide thanking someone for being there, thanking someone for their time, thankful sounds, soft, male voice	can hear the mouth full of the woman, as if she has too much saliva while speaking, robotic sound and constant sounds no particular changes in pitch	sounds like a dome or someone in a church chanting with the person is saying a group of people responding in unison and one person chanting with authority what to say	confident voice, reminds me of a Scottish accent due to the lowered intonation, it sounds very descriptive		
P.	Sound of my laptop. The sound of a bus outside my window	Scary sound form far away	Distorted sound	Waves	People talking about pictures	sounds like a lecture ?	Echo	Tunnel	sounds like someone speaking rly fast		
M.		feels like asthma. Short breaths. Higher pitch, lots of breaks, slight metallic feel	peak in a particular frequency, less balanced, single high point, dry	mouth full, and slight whistle/ breathy side, clearest words yet	even clearer, sounds like they were injured in the mouth/ had a tooth removed, phrasing direct, free of emotions	speaking with something hot in the mouth, slurry	high pitch beep, metallic, bitter?	crackle, layered, scrunchy. Separate from a voice	hissy, slight swoosh at the end of words	very clear and defined, detailed dry	spoken through a cardboard tube, s and f amplified into a harsh windy sound.
O.		female voice sounds hollow, slight buzzing noise underlying.	voice sounds a bit like it's slurring and like the mouth is not closing "properly"	sounds like mouth is rounding deliberately, more intelligible + lisp	clear voice, articulation quite deliberate	sounds like speaker is speaking around something in their mouth	slightly tinny, sounds like speaking through something (like a mask)	background crinkling.	stop sounds not really reverb - sounds a bit like voice is trying to emulate an old person (maybe with missing teeth?)	slightly nasal voice.	fricatives seem extra loud / drawn out so somewhat hissy, not comfortable to listen to
S.		the sound seems bothered by the person having their hand over their mouth	very muffled sound	the mouth of this young boy seems closed making an oval shape with their mouth, I feel their eyes are also slightly closed	a teacher speaking about grammar with a monotone voice but seems like the teeth were pinching the cheeks	big object in the boy's mouth	feels like a screen over the face of the speaker and voice bouncing back slightly	feels like typing/ scratching sounds in the background but the voice is clearer than the others	feels like this person has no teeth, or lips tightly pursed	very nasal sound as if she's pinching her nose, very heavy breathing	feels like the woman is speaking through a cone over her mouth to speak, clear pitch and happy voice
P.											

Appendix D:

IBM Workshop Google Form Responses

Do you work with voice in your work? Please describe in what ways...

2 responses

I am analysing interview transcripts based mostly on text but in the future also on speech.

yes - IVR & virtual assistants via voice channel (usually telephone)

Do you work with AI in your work? Please describe in what ways...

2 responses

I am an AI consultant, where I design AI solutions for specific clients and projects.

yes- Conversational AI

Do you work with sound or listening in your work? Please describe in what ways...

2 responses

I do not.

somewhat - listening as in, listening to entire customer calls, listening to prompts to decide which ones sound better (by the voice/TTS)

Was there something we did / talked about today that made you think differently about voice?

2 responses

Yes, identity of voices. It made me think about personalising the vocal side of virtual assistants. Also the idea that a single person may have multiple voices and the usage of different voices for different contexts.

I found the embodiment of voice really interesting as I hadn't really thought about that before (in terms of - personification of a virtual assistant, perhaps should also include how they look?)

How is what we've done / talked about today different from how you usually understand voice?

2 responses

It put more importance on the actual sound and it's characteristics as opposed to the words uttered.

I think I've not really considered the "noises" to be part of voice, or the entire audio experience

Is there anything we did / talked about today that you would be interested to implement into / explore in your work?

2 responses

Yes potentially designing custom voice identities for custom use cases.

not sure - potentially the personification aspect

What was something you did / found out today that was new or particularly interesting to you?

2 responses

That there exists AI that can quite accurately create faces based on speech.

freeform listening to sounds was new - I guess I hadn't purposefully done that (apart from perhaps in yoga /meditation practice), thinking about work!

Please add any other comments

2 responses

Good luck on the final six weeks!

I do think exploration of how a voice is perceived is really important when communicating with customers over voice channel/telephony, as that is the only thing they experience (as opposed to, say, a webchat where you'll also have the website presence, design)

Appendix E:

Public Engagements Since Commencing PhD Study - January 2024

2023

Speaker, *AI: Who's Looking After Me?*, Friday Lates: Machine Mythologies, Science Gallery, London

Exhibition, *Techne 1 Showcase Event*, ICA, London

Exhibition, *RCA Research Biennale 2023*, Copeland Gallery, Peckham

Performance, *What You Will Watch and Hear*, performed by Musarc, Lisson Gallery, London

Speaker, *Centre for Visual Cultures*, Royal Holloway

Writing, *Sounding Out!* Blog, Online

Speaker & Workshop, *Articulating Data Symposium*, University of Edinburgh

Speaker, *Uncanny Machines*, Scottish AI Summit with Johann Diedrick

Commission, *The New Real AI Art Commission: Uncanny Machines* with Johann Diedrick

2022

Solo Performer, *Can you hear it?* for Fani Parali, Cooke Latham Gallery, London

Guest Lecture, *Art and Design Department*, University of Leeds, Leeds

Publication, *Design Issues*. MIT Press

Peer Review, *openwork*, Columbia University

2021

Book Contribution, *Hausmusik Kollektiv*, edited by Claudia Molitor, Uniformbooks

Speaker, *Auraldiversities: Future Listening*. Goldsmiths University

Speaker, *Looking Back and Looking Forward*, Sound Practise and Research, City, University of London

Workshop, with Anja Borowicz-Richardson, *Giving Voice to Synthetic Sonics*, RCASU

Speaker, *Art/Thought/Sound: Knowing Through Sound*, School of Arts, UCP, Porto, Portugal

Exhibition, *Research Biennale*, RCA, Online

2020

Interview, *TECHnique podcast: Episode 44*, Online

Solo Performer, *Aonyx And Drepan, The Minders Of The Warm* by Fani Parali, Southwark Park Galleries

Exhibition, *Attune*, Research and Waves, Online

Solo Performer, *Repositorium* by Nestor Pestana, Scrolling The Arcane, Planetarium, Porto, Portugal

Exhibition, *Empathy Loading*, Furtherfield Gallery x Curating Contemporary Art, RCA. Online.

Guest Lecture, *Masters Research Students*, Royal College of Art, London

Guest Lecture, *Art and Design Department*, University of Leeds, Leeds

2019

Guest Lecture, *Music Department*, City, University of London

Writing, *Ada Lovelace Imagining the Analytical Engine*, Programme Notes, Barbican Centre, London

Guest Lecture, *Art and Design Department*, Richmond University, London

Speaker, *Human-Data Interaction workshop on Music and AI*, Somerset House, London

Speaker, *SPARC (Sound Practise and Research at City, University of London) Symposium*, London

Workshop, Summer School, Tate Modern, London
Exhibition / Installation, *Le Marteau Sans Maître Concert*, Musarc, Whitechapel Bell Foundry
Speaker, *The (Un)Sound Barrier Symposium*, Royal College of Art, London
Solo Performer, *The Terrace of Lungs* by Fani Parali, Zabłudowicz Collection, London
Workshop, *Life Rewired Hub*, Barbican Centre, London.
Solo Performer, *Angels like Buildings* by Fani Parali, Her Voice, ICA, London
Solo Performer, *Ecstasies* by Marguerite Humeau, Kunstverein, Hamburg
Performer, with East Anglia Records, *Ambit Magazine 235 Launch*, Tate Modern
Solo Performer, *Angels like Buildings* by Fani Parali, Assembly Point, London
Screening, *Four Words Future*, All 4, Channel Four Television

2018

Performer, with Musarc, *London Contemporary Music Festival*, London
Performer, with East Anglia Records, *Oral Rinse 5*, SET, London

References

- Abbas-Nazari, A. (2014) *Across the sonic border*. 2014. Speculative Voicing. <https://speculativevoicing.co.uk/Across-The-Sonic-Border> [Accessed: 12 July 2022].
- Abbas-Nazari, A. (2020) *Acoustic Ecology of an AI System*. [Online]. 2020. Attune. Available from: <https://attune.researchandwaves.net/acoustic-ecology-of-an-ai-system.html> [Accessed: 16 August 2021].
- Abbas-Nazari, A. (2022) *Speculative Voicing*. 2022. <https://speculativevoicing.co.uk/> [Accessed: 18 July 2022].
- Abbas-Nazari, A. (2023) Beyond the every day: vocal potential in AI mediated communication. *Sounding Out!*. <https://soundstudiesblog.com/2023/05/22/vocal-potential-in-ai-mediated-communication/>.
- Abbas-Nazari, A. (n.d.) @speculative_voicing • Instagram photos and videos. https://www.instagram.com/speculative_voicing/ [Accessed: 15 August 2022].
- Abbas-Nazari, A., Borowicz-Richardson, A. (2021) Workshop materials, 2021. <https://drive.google.com/drive/u/0/folders/1cWMgLrWuXGIgXyYYJoPgFuGXYqMXuNna> [Accessed: 31 August 2022].
- Abercrombie, G., Curry, A.C., Pandya, M. & Rieser, V. (2021) Alexa, Google, Siri: what are your pronouns? Gender and anthropomorphism in the design and perception of conversational assistants. *arXiv:2106.02578 [cs]*. [Online]. Available from: <http://arxiv.org/abs/2106.02578> [Accessed: 22 June 2021].
- Aboutmatch (n.d.) *The role of accent in UK call centres*. Aboutmatch. [Online]. Available from: <https://www.aboutmatch.co.uk:443/Aboutmatch/Forms/Article/The%20role%20of%20accent%20in%20UK%20call%20centres.aspx> [Accessed: 17 September 2021]
- Abu Hamdan, L. (2018) *Aural contract: investigations at the threshold of audibility*. PhD thesis, Goldsmiths, University of London. [Online]. Available from: <http://research.gold.ac.uk/id/eprint/23293/> [Accessed: 30 August 2021].
- Adams, T.E., Ellis, C. & Jones, S.H. (2017) Autoethnography. In: *The International Encyclopedia of Communication Research Methods*. Chichester: John Wiley,. pp. 1–11. doi:[10.1002/9781118901731.iecrm0011](https://doi.org/10.1002/9781118901731.iecrm0011).
- Ahmed, A.A., Kok, B., Howard, C. & Still, K. (2021) Online community-based design of free and open source software for transgender voice training. In *Proceedings of the ACM on*

Human-Computer Interaction. 4 (CSCW3), 258:1-258:27. doi:[10.1145/3434167](https://doi.org/10.1145/3434167).

Alexa Developer (n.d.) *Speech synthesis markup language (SSML) Reference* | *Alexa Skills Kit*. Amazon (Alexa). <https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html> [Accessed: 8 December 2022].

Amacher, M. (1980 –) *Intelligent life*.

Amaro, R. (2022) *The Black technical object: on machine learning and the aspiration of Black being*. Volume 2. Berlin: Sternberg Press.

Amaro, R. (2019) As if. [Online]. *E-Flux*, February. Available from: <https://www.e-flux.com/architecture/becoming-digital/248073/as-if/> [Accessed: 11 May 2021].

Amazon (2023) *Amazon introduces four all-new Echo devices; sales of Alexa-enabled devices surpass half a billion*. 17 May. Amazon Press Center. <https://press.aboutamazon.com/2023/5/amazon-introduces-four-all-new-echo-devices-sales-of-alexa-enabled-devices-surpass-half-a-billion> [Accessed: 26 September 2023].

Amazon (n.d.) *What Is Alexa Voice ID? - Amazon Customer Service*. <https://www.amazon.com/gp/help/customer/display.html?nodeId=GYCXY2AB2QWZT2X> [Accessed: 26 July 2022].

Anderson, L. (1982) *O Superman*. Warner Bros.

Aneesh, A. (2015) *Neutral Accent: How Language, Labor, and Life Become Global*. Durham, NC; London: Duke University Press.

Anon (n.d.) *Conversational AI: What is conversational AI?* [Online]. Interactions. Available from: <https://www.interactions.com/conversational-ai/> [Accessed: 9 August 2021].

Articulating Data (2023) *Articulating Data*. 2023. <https://articulatingdata.com/> [Accessed: 1 August 2023].

Art Viewer (2019). *Marguerite Humeau at Kunstverein Hamburg*. Art Viewer, March 31 <https://artviewer.org/marguerite-humeau-at-kunstverein-hamburg/>.

Attune (n.d.) *Attune*. <https://attune.researchandwaves.net/> [Accessed: 26 September 2023].

Audacity (n.d.) *Audacity*® | Free, open source, cross-platform audio software for multi-track recording and editing. <https://www.audacityteam.org/> [Accessed: 22 November 2023].

audEERING (2021) *audEERING Homepage*. 4 September 2021. <https://>

www.audeering.com/ [Accessed: 1 August 2022].

Auger, J. (2013) Speculative Design: Crafting the Speculation. *Digital Creativity*. 24. doi:[10.1080/14626268.2013.767276](https://doi.org/10.1080/14626268.2013.767276).

Baird, A., et al. (2017) Perception of paralinguistic traits in synthesized voices. In *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, 23 August 2017, pp. 1–5. doi:[10.1145/3123514.3123528](https://doi.org/10.1145/3123514.3123528).

Bakhtiari, K. (2022) Gen-Z demand racial justice, not just diversity, equity and inclusion from brands. *Forbes*, 5 June <https://www.forbes.com/sites/kianbakhtiari/2022/06/05/gen-z-demand-racial-justice-not-just-diversity-equity-and-inclusion-from-brands/> [Accessed: 19 December 2023].

Barad, K. (2007) *Meeting the universe halfway: quantum physics and the entanglement of matter and meaning*. Durham, NC: Duke University Press.

Barbican (2019) *Squish Space*. Barbican. 25 April. <https://www.barbican.org.uk/whats-on/2019/event/life-rewired-hub-residency-squish-space> [Accessed: 12 April 2023].

Barbican (2019 b) *Trevor Paglen: From ‘Apple’ to ‘Anomaly’*. Barbican. Available from <https://www.barbican.org.uk/whats-on/2019/event/trevor-paglen-from-apple-to-anomaly> [Accessed: 9 May 2022].

Barthes, R. (1977) The grain of the voice. In *Image, music, text: essays*. London: Fontana, pp. 179–189.

Baugh, J. (2002) Linguistic profiling. In: A. Ball, S. Makoni, G. Smitherman, A.K. Spears, & F. by N. wa Thiong’o (eds.). *Black linguistics: language, society and politics in Africa and the Americas*. London: Routledge., pp. 155–168.

BBC (2021) *Alexa tells 10-year-old girl to touch live plug with penny*. BBC News, 28 December. <https://www.bbc.com/news/technology-59810383>.

Beckett, L. & Paul, K. (2023) ‘Bargaining for our very existence’: why the battle over AI is being fought in Hollywood. *Guardian*, 22 July. <https://www.theguardian.com/technology/2023/jul/22/sag-aftra-wga-strike-artificial-intelligence>.

Behar, K. (2018) Personalities without people / anonymous autonomous. Lecture. What’s Next: Distributed Technology and the Technosphere. RCA, London.

Bender, E.M., et al. (2021) On the dangers of stochastic parrots: can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and*

Transparency. FAccT '21. [Online]. 3 March 2021 New York, NY, USA, Association for Computing Machinery, pp. 610–623. Available from: doi:[10.1145/3442188.3445922](https://doi.org/10.1145/3442188.3445922) [Accessed: 9 August 2021].

Bennett, J. (2009) *Vibrant matter: a political ecology of things*. Durham, NC: Duke University Press.

Benjamin, R. (2019) *Race after technology: abolitionist tools for the New Jim Code*. Cambridge: Polity Press.

Berardi, F. (2018) *Breathing: chaos and poetry*. Semiotext(e) intervention series 26. South Pasadena, CA, Semiotext(e).

Berberian, C. (1966) *Stripsody*. Available from: <https://www.youtube.com/watch?v=0dNLAhL46xM>.

Birhane, A. (2021) *The impossibility of automating ambiguity*. Artificial Life. [Online] 27 (1), 44–61. doi:[10.1162/artl_a_00336](https://doi.org/10.1162/artl_a_00336).

Birhane, A. (2022) *Automating ambiguity: challenges and pitfalls of artificial intelligence*. PhD thesis, University College Dublin. <http://arxiv.org/abs/2206.04179>.

Birhane, A., Prabhu, V.U. & Kahembwe, E. (2021) Multimodal datasets: misogyny, pornography, and malignant stereotypes. *arXiv:2110.01963 [cs]*. <http://arxiv.org/abs/2110.01963>.

Black, A.W., Zen, H. & Tokuda, K. (2007) Statistical parametric speech synthesis. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07. [Online]. April 2007 pp. IV-1229-IV-1232. Available from: doi:[10.1109/ICASSP.2007.367298](https://doi.org/10.1109/ICASSP.2007.367298).

Blessner, B. & Salter, L.-R. (2009) *Spaces speak, are you listening? experiencing aural architecture*. Cambridge, MA: MIT Press.

Blue, L., et al. (2022) Who are you (I really wanna know)? Detecting audio deepfakes through vocal tract reconstruction. In: *31st USENIX Security Symposium (USENIX Security 22)*, pp. 2691–2708.

Boesch, G. (2021) *What is adversarial machine learning?* viso.ai. <https://viso.ai/deep-learning/adversarial-machine-learning/>

Bowden, C. (2016) *Calls of duty*. Calsbo. <https://calsbo.com/calls-of-duty/>.

Bratton, B.H. (2016) *The stack: on software and sovereignty*. Cambridge, MA: MIT Press.

Brown, A. (2021) *Voice recognition is awesome, but how did it get so good?* MUO, 2 October.

<https://www.makeuseof.com/voice-recognition-improve/> [Accessed: 22 August 2022].

Bruder, J. (2020) Alexa's body: what the interface obscures and how design could help us see. In: Claudia Mareis (ed.). *Design struggles: intersecting histories, pedagogies, and perspectives*. Amsterdam: Valiz, pp. 283–297.

Buolamwini, J. & Gebru, T. (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. 21 January 2018. PMLR, pp. 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>.

Burton, M. & Nitta, M. (2019) *New organs of creation*. Burton Nitta. <https://www.burtonnitta.co.uk/NewOrgansOfCreation.html#> [Accessed: 2 January 2024].

Camp, I. (2020) *Behind Shirley*. <https://vimeo.com/464136359>.

Candela, E. & de Visscher, E. (2023) Learning from 'The Sounding Object': sound design in the critical reimagining of museum object narratives. *Design Issues*. 39 (2), 57–71. doi:[10.1162/desi_a_00717](https://doi.org/10.1162/desi_a_00717).

Candy, L. (2006) *Practice-based research: a guide*. [Online]. 2006. Creativity & Cognition Studios. Available from: <https://www.creativityandcognition.com/resources/PBR%20Guide-1.1-2006.pdf> [Accessed: 10 June 2021].

Carnegie, M. (2022) *Gen Z: How young people are changing activism*. BBC (Worklife). <https://www.bbc.com/worklife/article/20220803-gen-z-how-young-people-are-changing-activism> [Accessed: 24 November 2023].

Carnegie Mellon University (n.d.) *Festvox: CMU_ARCTIC Databases*. http://festvox.org/cmu_arctic/ [Accessed: 14 August 2021].

Carpenter, E., & McLuhan, M. (1960). Acoustic space. In E. Carpenter & M. McLuhan (eds.), *Explorations in communication*. Boston, MA: Beacon Press, pp. 65-70.

Cavarero, A. (2005) *For more than one voice: toward a philosophy of vocal expression*. Stanford, CA: Stanford University Press.

Cave, N. (1992 -) *Soundsuits*.

Ceraso, S. (2022) *Voice as ecology: voice donation, materiality, identity*. Sounding Out!. <https://soundstudiesblog.com/2022/09/06/voice-as-ecology-voice-donation-materiality-identity/>.

Chan, W. (2022) The AI startup erasing call center worker accents: is it fighting bias – or perpetuating it? *Guardian*. 24 August. <https://www.theguardian.com/technology/2022/>

[aug/23/voice-accent-technology-call-center-white-american.](#)

Chion, M. (1994) *Audio-vision: sound on screen*. New York: Columbia University Press.

Cimini, A. (2019) In your head: notes on Maryanne Amacher's *Intelligent Life*. In: *Blank Forms 4: Intelligent Life*. New York, Blank Forms Editions. pp. 151–196.

Clearspeed (n.d.) *Clearspeed Product Data Sheet*. [Online]. Clearspeed. Available from: https://www.clearspeed.com/wp-content/uploads/2020/08/Clearspeed-Product-Data-Sheet_081120.pdf [Accessed: 12 March 2021].

Collins English Dictionary (2023) Voice synthesiser. *Collins English Dictionary*. <https://www.collinsdictionary.com/dictionary/english/voice-synthesizer> [Accessed: 3 December 2023].

Connor, S. (2001) Satan and Sybil: talk, possession, and dissociation. In: S.I. Salamensky (ed.). *Talk, talk, talk: the cultural life of everyday conversation*. New York: Routledge, pp. 163–180.

Connor, S. (2015) *Choralities - Steven Connor*. [Online]. Steven Connor. Available from: <http://stevenconnor.com/choralities.html> [Accessed: 4 July 2022].

Copenhagen Pride, Virtue, Equal AI, Koalition Interactive, et al. (n.d.) *Meet Q. The first genderless voice*. [Online]. Genderless Voice. Available from: <https://www.genderlessvoice.com> [Accessed: 5 September 2021].

Costanza-Chock, S. (2018) Design justice, A.I., and escape from the matrix of domination. *Journal of Design and Science*. doi:[10.21428/96c8d426](https://doi.org/10.21428/96c8d426).

Cox, T. (2019) *Now you're talking: human conversation from the neanderthals to artificial intelligence*. London: Vintage.

Crabbé, J. & van der Schaar, M. (2022) Label-free explainability for unsupervised models. *arXiv* 2203.01928. doi:[10.48550/arXiv.2203.01928](https://doi.org/10.48550/arXiv.2203.01928).

Crawford, K. (2021) *Atlas of AI: power, politics, and the planetary costs of artificial intelligence*. New Haven, CT: Yale University Press.

Crawford, K. & Joler, V. (2018) *Anatomy of an AI system*. [Online]. Available from: <https://anatomyof.ai/> [Accessed: 26th March 2020]

Crenshaw, K. (1989) Demarginalizing the intersection of race and sex: a Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*. 139. <https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>.

Cummins, F. (2020) The territory between speech and song: a joint speech perspective. *Music Perception*. 37 (4), 347–358. doi:[10.1525/mp.2020.37.4.347](https://doi.org/10.1525/mp.2020.37.4.347).

Cycling '74 (n.d.) *What is Max?* Cycling '74. <https://cycling74.com/products/max> [Accessed: 12 September 2023].

Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G.J., Durand, F. & Freeman, W.T. (2014) The visual microphone: passive recovery of sound from video. *ACM Transactions on Graphics*. 33 (4), 79:1-79:10. doi:[10.1145/2601097.2601119](https://doi.org/10.1145/2601097.2601119).

Debatty, R. (2011) *The rebirth of prehistoric creatures*. We Make Money Not Art. https://we-make-money-not-art.com/back_here_below_formidable/.

Dery, M. (1993) Black to the future: interviews with Samuel R. Delany, Greg Tate, and Tricia Rose/Mark Dery. In: *Flame wars: the discourse of cyberculture*. Durham, NC: Duke University Press. pp. 179–222.

Derrida, J. (1998) The separation of speech and song. In: *Of Grammatology*. Corrected edition. Baltimore, MD: Johns Hopkins University Press, pp. 195–200.

Descript (n.d) *Overdub: natural sounding text-to-speech*. Descript. Available from: <https://www.descript.com/overdub> (Accessed: 7 September 2023).

Deutsch, D. (1995) *Speech-to-song illusion*. Diana Deutsch. <https://deutsch.ucsd.edu/psychology/pages.php?i=212> [Accessed: 5 September 2022].

Diedrick, J. *Dark matters*. (2021) <https://darkmatters.ml/>

Disley, M. & Khan, M. (2021) *UNSOUND DAY 3: Speculative Voices and Machine Learning*. Unsound Festival. [YouTube video]. <https://www.youtube.com/watch?v=5kcXnRE0830>.

Dolar, M. (2006) *A Voice and nothing more*. Cambridge, MA: MIT Press.

Du Bois, W. E. B. (1903) *The souls of Black folk*. Chicago, IL: A. C. McClung.

Dudley, H. (1940) The carrier nature of speech. *The Bell System Technical Journal*. [Online] 19 (4), 495–515. Available from: doi:[10.1002/j.1538-7305.1940.tb00843.x](https://doi.org/10.1002/j.1538-7305.1940.tb00843.x).

Dunne, A. (2009) One million little utopias. In: Onkar Kular (ed.). *Accept no other imitations*. London: Royal College of Art, Design Interactions.

Dunne, T., Raby, F. (2014) *Speculative everything: design, fiction, and social dreaming*. Cambridge, MA: MIT Press.

Edwards, B. (2023) Microsoft's new AI can simulate anyone's voice with 3 seconds of

audio. 9 January. *Ars Technica*. <https://arstechnica.com/information-technology/2023/01/microsofts-new-ai-can-simulate-anyones-voice-with-3-seconds-of-audio/> [Accessed: 16 January 2023].

Eidsheim, N.S. (2012) Voice as action: towards a model for analyzing the dynamic construction of racialized voice. *Current Musicology*. 93. doi:[10.7916/cm.v0i93.5218](https://doi.org/10.7916/cm.v0i93.5218).

Eidsheim, N.S. (2015) *Sensing sound: singing and listening as vibrational practice*. Durham, NC: Duke University Press.

Eidsheim, N.S. (2019) *The race of sound*. Durham, NC: Duke University Press.

Elmjouie, Y. (2014) Alone again, naturally: women singing in Iran. *Guardian*. [Online] 29 August. Available from: <https://www.theguardian.com/world/iran-blog/2014/aug/29/women-singing-islamic-republic-iran> [Accessed: 31 October 2021].

Empathy Loading (2020) *Empathy Loading*. 2020. <https://empathyloading.com/> [Accessed: 12 April 2023].

Empathy Loading (2020 b) *Empathy Loading (@empathyloading)* [Twitter]. 15 June 2020. <https://twitter.com/empathyloading> [Accessed: 25 April 2023].

Encyclopedia Britannica (n.d.) Polyphony [Online]. *Encyclopedia Britannica*. Available from: <https://www.britannica.com/art/polyphony-music> [Accessed: 30 October 2021].

Ephrat, A., Mosseri, I., Lang, O., Dekel, T., et al. (2018) Looking to listen at the cocktail party: a speaker-independent audio-visual model for speech separation. *arXiv:1804.03619* [cs, eess]. [Online] Available from: doi:[10.1145/3197517.3201357](https://doi.org/10.1145/3197517.3201357) [Accessed: 5 August 2021].

Eshun, K. (1998) *More brilliant than the sun: adventures in sonic fiction*. London: Quartet Books.

Eubanks, V. (2018) *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St Martin's Press.

Feldman, J. (2016) "The problem of the adjective": affective computing of the speaking voice. *Transposition. Musique et Sciences Sociales*. (6). doi:[10.4000/transposition.1640](https://doi.org/10.4000/transposition.1640).

Field Studies (2017) *Field Studies 2017: Listening after Pauline Oliveros*. Leeds, 12–15 Oct. Programme 2017. <http://field-studies.org/programme-2017/> [Accessed: 10 August 2023].

Franinovic, K. & Serafin, S. (2013) *Sonic interaction design*. Cambridge, MA: MIT Press.

Fraenkel Gallery (2021) *Christian Marclay*. Fraenkel Gallery. <https://fraenkelgallery.com/exhibitions/christian-marclay> [Accessed: 30 November 2022].

Gaviny (2011) *Underground Resistance - The Force - Album-Death Star - 1992*. [YouTube video] <https://www.youtube.com/watch?v=XIXcOkDouXA>.

Gilbert, R.Y. (1911) Vocal fingerprints, *San Francisco Call*, July 16.

Goodfellow, I., Bengio, Y. & Courville, A. (2016) *Deep learning*. Cambridge, MA: MIT Press.

Goodfellow, I., et al. (2014) Generative adversarial networks. *Advances in Neural Information Processing Systems*. 3. doi:[10.1145/3422622](https://doi.org/10.1145/3422622).

gov.uk (2023) *UK Population by Ethnicity*. 31 March 2023. <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest/> [Accessed: 25 January 2024].

Gow, G. (2001) Spatial metaphor in the work of Marshall McLuhan. *Canadian Journal of Communication*. 26. Available from: doi:[10.22230/cjc.2001v26n4a1254](https://doi.org/10.22230/cjc.2001v26n4a1254).

Hao, K. (2020) We read the paper that forced Timnit Gebru out of Google. Here's what it says. *MIT Technology Review*. 4 December. <https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/> [Accessed: 5 September 2022].

Hao, K. (2020 b) The two-year fight to stop Amazon from selling face recognition to the police. 12 June 2020. *MIT Technology Review*. 12 June. <https://www.technologyreview.com/2020/06/12/1003482/amazon-stopped-selling-police-face-recognition-fight/> [Accessed: 27 November 2022].

Hazirbas, C., et al. (2021) Towards measuring fairness in AI: the casual conversations dataset. *arXiv:2104.02821 [cs]*. <http://arxiv.org/abs/2104.02821>.

He, N. (2019) *Robots shouldn't sound human:...* A MAZE: Berlin 2019: [Online]. Available from: <https://amazeberlin2019.sched.com/event/NE3y/nicole-he-robots-shouldnt-sound-human-the-aesthetics-of-the-computer-voice-in-art-and-games> [Accessed: 5 November 2020].

Henriques, J. (2011) *Sonic bodies: reggae sound systems, performance techniques, and ways of knowing*. New York: Continuum.

Herndon, H. (2021). *Holly+*. *Holly+— Mirror* <https://holly.mirror.xyz/54ds2liOnvthjGFkokFCoal4EabytH9xjAYy1irHy94> [Accessed: 1 November 2021].

Hoffmann, A. L. (2018) Data violence and how bad engineering choices can damage Society. [Online]. *Medium*. Available from: <https://medium.com/s/story/data-violence->

[and-how-bad-engineering-choices-can-damage-society-39e44150e1d4](#) [Accessed: 29th April 2020]

Holly+ (n.d.) *Genesis - Holly+ Raw Singing Model I*. <https://holly-plus-auction.vercel.app> [Accessed: 19 August 2022].

Hoo, W.S. (2012) Nick Cave performs at the U.S. State Department's Art in Embassies 50th anniversary celebration. *Washington Post*, 28 November. https://www.washingtonpost.com/blogs/arts-post/post/soundsuits-sculptor-nick-cave-performs-at-the-us-state-departments-art-in-embassies-50th-anniversary-celebration/2012/11/28/ab97c740-39a1-11e2-a263-f0ebffed2f15_blog.html.

Hutson, M. (2021) Who should stop unethical A.I.? *The New Yorker*, February 15. <https://www.newyorker.com/tech/annals-of-technology/who-should-stop-unethical-ai>.

Humeau, M. (2011) *Back, Here, Below, Formidable*.

Hunt, A.J. & Black, A.W. (1996) Unit selection in a concatenative speech synthesis system using a large speech database. In: *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. [Online]. 1996 Atlanta, GA, USA, IEEE. pp. 373–376. Available from: doi:[10.1109/ICASSP.1996.541110](https://doi.org/10.1109/ICASSP.1996.541110) [Accessed: 18 August 2021].

IBM (n.d.) *IBM Voice Surveillance*. IBM. <https://www.ibm.com/docs/en/siffs/2.0.1?topic=services-voice-surveillance> [Accessed: 1 August 2022].

IBM Watson (2021) *IBM Watson*. IBM. <https://www.ibm.com/uk-en/watson> [Accessed: 8 August 2022].

Ilde, D. (2007) *Listening and voice: phenomenologies of sound*. 2nd edition. New York: State University of New York Press.

Intel (n.d.) *Zoom uses AI to improve virtual meetings*. Intel. <https://www.intel.com/content/www/us/en/customer-spotlight/stories/zoom-customer-story.html> [Accessed: 8 May 2022].

James, R. (2019) *The sonic episteme: acoustic resonance, neoliberalism, and biopolitics*. Durham, NC: Duke University Press.

Jin, H. & Wang, S. (2018) *Voice-based determination of physical and emotional characteristics of users*. US 10,096,319 (Patent) [Online]. <https://patentimages.storage.googleapis.com/f6/a2/36/d99e36720ad953/US10096319.pdf>

Jasanoff, S. & Kim, S.-H. (2009) Containing the atom: sociotechnical imaginaries and nuclear power in the United States and South Korea. *Minerva*, 47(2), 119–146.

Jasanoff, S. (2015) Future imperfect: science, technology, and the imaginations of modernity. In: S. Jasanoff & S.-H. Kim (eds.). *Dreamscapes of modernity*. Chicago, IL: University of Chicago Press. pp.1–33.

Kang, E.B. (2022) Biometric imaginaries: formatting voice, body, identity to data. *Social Studies of Science*. 52(4). doi:[10.1177/03063127221079599](https://doi.org/10.1177/03063127221079599).

Karantonis, P. & Verstraete, P. (2014) Introduction/Overture. In: *Cathy Berberian: pioneer of contemporary vocalty*. London: Routledge, pp. 3-18

Karikis, M (2017) Mikhail Karikis 'The Voice as Sculpture' (Interview). <https://vimeo.com/239862208>.

Keyes, O. (2019) Counting the countless: why data science is a profound threat for queer people. [Online]. *Real Life*. Available from: <https://reallifemag.com/counting-the-countless/> [Accessed: 17 November 2020].

Khan, M., Disley M. (2023) *Not I*

Kidel, S. (2018) *Voice Recognition DoS Attack*. Bandcamp. <https://samkidel.bandcamp.com/>.

Kitamura, T. & Ohtani, K. (2015) Non-contact measurement of facial surface vibration patterns during singing by scanning laser Doppler vibrometer. *Frontiers in Psychology*. 6. <https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01682>.

Koenecke, A., Nam, A., Lake, E., Nudell, J., et al. (2020) Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*. 117 (14), 7684–7689. Available from: doi:[10.1073/pnas.1915768117](https://doi.org/10.1073/pnas.1915768117)

Krejci, J. (2018) Giving it her voice. *Form Magazine*. 279 (Sept / Oct),.46–53.

Lau, J., Zimmerman, B. & Schaub, F. (2018) Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. In *Proceedings of the ACM on Human-Computer Interaction*. [Online] 2 (CSCW), 102:1-102:31. Available from: doi:[10.1145/3274371](https://doi.org/10.1145/3274371).

Lemoine, B. (2022) Is LaMDA Sentient? An interview. *Medium*. June 11. <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917>.

Levanon, Y. & Lossos, L. (2011) *System for indicating emotional attitudes through intonation analysis and methods thereof*. US 8,078,470 B2 (Patent). <https://patentimages.storage.googleapis.com/50/43/54/05743cf3e41181/US8078470.pdf>

Leviathan, Y. & Matias, Y. (2018) *Google Duplex: An AI System for Accomplishing Real-World*

Tasks Over the Phone. <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>.

Li, X. & Mills, M. (2019) Vocal features: from voice identification to speech recognition by machine. *Technology and Culture*. 60, S129–S160. Available from: doi:[10.1353/tech.2019.0066](https://doi.org/10.1353/tech.2019.0066).

Lombog, M. (2023) 'Quickly build an Alexa skill to talk with ChatGPT', *Medium*, 22 March. Available from: <https://medium.com/geekculture/quickly-build-an-alexa-skill-to-talk-with-chatgpt-7feaeed48b0> (Accessed: 7 September 2023).

Lorde, A. (1984) The master's tools will never dismantle the master's house. In: *Sister Outsider*. Berkeley, CA: Crossing Press, pp. 110–113.

Lorde, A. (1984 b) The Transformation of Silence into Language and Action. In: *Sister Outsider*. California, USA, The Crossing Press. pp. 40–44.

Luck, N. & Musarc Choir (2013) *Namesaying*. Soundcloud. Available from: <https://soundcloud.com/musarc/musarc-neil-luck-namesaying-2013>.

Madeleine, A. (2014) Artist Nick Cave's wild and whimsical soundsuits – in pictures. *Guardian*, 13 November. <http://www.theguardian.com/artanddesign/gallery/2014/nov/13/artist-nick-caves-soundsuits-pictures>.

Mager, A. & Katzenbach, C. (2021) Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society*. 23 (2), 223–236. doi:[10.1177/1461444820929321](https://doi.org/10.1177/1461444820929321).

Marclay, C. (2020) *No!*

Martins, L. (2014) Privilege and oppression: towards a feminist speculative design. In Lim, Y., et al. (eds.), *Design's Big Debates - DRS International Conference 2014*, 16-19 June, Umeå, Sweden. Available from: <https://dl.designresearchsociety.org/drs-conference-papers/drs2014/researchpapers/75/>

Mashable Deals (n.d.) *Google's AI assistant can now make real phone calls*. [YouTube video]. https://www.youtube.com/watch?v=JvbHu_bVa_g [Accessed: 22 June 2021].

Meizel, K. (2020) *Multivocality: singing on the borders of identity*. New York, NY: Oxford University Press.

Merriam-Webster Dictionary (n.d.) Profiling. [Online]. *Merriam-Webster Dictionary*. Available from: <https://www.merriam-webster.com/dictionary/profiling> [Accessed: 30 October 2021].

Montgomery, E.P. (n.d.) *An unresolved mapping of speculative design V 2.0*. EPMID.

<https://www.epmid.com/projects/Mapping-Speculative-Design> [Accessed: 20 September 2022].

Moreland, Q. (2019) Meara O'Reilly: *Hockets for Two Voices* EP. Pitchfork. <https://pitchfork.com/reviews/albums/meara-oreilly-hockets-for-two-voices-ep/> [Accessed: 21 September 2023].

Mori, M., MacDorman, K. & Kageki, N. (2012) The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*. 19 (2), 98–100. Available from: doi:[10.1109/MRA.2012.2192811](https://doi.org/10.1109/MRA.2012.2192811).

Moynihan, H.S., Qayyah (2020) Companies like Amazon may give devices like Alexa female voices to make them seem 'caring'. [Online]. *Business Insider*, April 5. Available from: <https://www.businessinsider.com/theres-psychological-reason-why-amazon-gave-alexa-a-female-voice-2018-9> [Accessed: 31 October 2021].

Mozilla (n.d.) *Mozilla Common Voice*. <https://commonvoice.mozilla.org/> [Accessed: 15 August 2023].

mturk (n.d.) *Amazon Mechanical Turk*. <https://www.mturk.com/worker/help> [Accessed: 7 January 2024].

Mulder, J. & Van Leeuwen, T. (2019) Speech, sound, technology. In: M. Grimshaw-Aagaard, M. Walther-Hansen and M. Knakkegaard (eds.) *The Oxford handbook of sound and imagination*, Volume 1. Oxford: Oxford University Press, pp. 473–492.

Müller, C. (2007) *Speaker classification II: selected projects*. Berlin; Heidelberg: Springer-Verlag.

multi'vocal collective (2021) The generation of a [multi'vocal] voice. *Seismograf*. [Online]. Available from: <https://seismograf.org/node/19502> [Accessed: 28 April 2021].

Musarc (n.d.) *Musarc*. <https://musarc.org/> [Accessed: 16 November 2022].

Musique Du Burundi. (1968) France: Ocora Records.

Nagrani, A., Albanie, S. & Zisserman, A. (2018) Seeing voices and hearing faces: cross-modal biometric matching. *arXiv:1804.00326 [cs]*. <http://arxiv.org/abs/1804.00326>.

Nass, C.I. & Brave, S. (2005) *Wired for speech: how voice activates and advances the human-computer relationship*. Cambridge, MA; London: MIT Press.

National Cyber Security Centre (2019) *Speaker recognition*. <https://www.ncsc.gov.uk/collection/biometrics/speaker-recognition> [Accessed: 5 September 2023].

Noble, S.U. (2018) *Algorithms of oppression: how search engines reinforce racism*. New York, NY: New York University Press.

Nygaard, K. (1990) The origins of the Scandinavian school, why and how. In *Proceedings of PDC '90 : Participatory Design Conference*, Seattle, March 31-April 1.

Oh, T.-H., Dekel, T., Kim, C., Mosseri, I., et al. (2019) Speech2Face: Learning the Face Behind a Voice. *arXiv:1905.09773* [cs]. [Online] Available from: <http://arxiv.org/abs/1905.09773> [Accessed: 5 November 2020].

Oliveira, P. (2016). Design at the Earview: Decolonizing Speculative Design through Sonic Fiction. *Design Issues*. 32 (2). doi: 10.1162/DESI_a_00381

Oliveira, P. (n.d.) PEDRO OLIVEIRA – DESMONTE. <https://oliveira.work/desmonte/> [Accessed: 13 August 2022].

Oliveros, P. (1971) *Ear Piece*.

Oliveros, P. (1974) *Your Voice*.

Oliveros, P. (1979) *Imaginary Meditations*.

Oliveros, P. (1985) *Sex Change*.

Oliveros, P. (2005) *Deep listening: a composer's sound practice*. New York, NY: iUniverse.

Oliveros, P. (2013) *Anthology of text scores*. Kingston, NY: Deep Listening Publications.

Oliveros, P. (2015) The difference between hearing and listening. Pauline Oliveros. *TEDx Indianapolis*. [YouTube video]. https://www.youtube.com/watch?v=_QHfOuRrJB8.

Oliveros, P. (n.d) *All or Nothing*

Oord, A. van den & Dieleman, S. (2016) *WaveNet: a generative model for raw audio*. Google DeepMind. <https://www.deepmind.com/blog/wavenet-a-generative-model-for-raw-audio> [Accessed: 7 August 2023].

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., et al. (2016) WaveNet: a generative model for raw audio. *arXiv:1609.03499* [cs]. [Online] Available from: <http://arxiv.org/abs/1609.03499> [Accessed: 7 October 2020].

OpenAI (2022) *DALL·E 2*. OpenAI. <https://openai.com/dall-e-2/> [Accessed: 5 September 2022].

OpenAI (2023) *Introducing ChatGPT*. Open AI. <https://openai.com/blog/chatgpt>

[Accessed: 20 September 2023].

Opiah, A. (2020) *UK music producer simulates illegal electronic rave in Google Data Centre*. [Online]. BroadGroup. Available from: <https://www.broad-group.com/data/news/documents/b1m2y1h8w8xqds/uk-music-producer-simulates-illegal-electronic-rave-in-google-data-centre> [Accessed: 3 September 2021].

O'Reilly, M. (2019) *Hockets for Two Voices*. <https://mearaoreilly.bandcamp.com/album/hockets-for-two-voices-ep>.

O'Reilly, M. *Illusion Songs*. [Online] Available from: <https://illusionsongs.tumblr.com/> [Accessed: 18th December 2019]

Owens, G. (2023) “Hey Google, talk like Issa”: Black voiced digital assistants and the reshaping of racial labor. *Sounding Out!*. <https://soundstudiesblog.com/2023/06/05/google-talk-like-issa-black-voiced-digital-assistants-and-the-reshaping-of-racial-labor/>.

Oxford Dictionary (n.d.) Normative. *Oxford Dictionary*. <https://www.oed.com/search/dictionary/?scope=Entries&q=normative> [Accessed: 30 September 2023].

Oxford Reference (n.d.) Ocularcentrism. *Oxford Reference*. doi:[10.1093/oi/authority.20110803100245338](https://doi.org/10.1093/oi/authority.20110803100245338) [Accessed: 21 September 2023].

Paglen, T. & Downey, A. (2020) Algorithmic anxieties: Trevor Paglen in conversation with Anthony Downey. *Digital War*. 1 (1), 18–28. doi:[10.1057/s42984-020-00001-2](https://doi.org/10.1057/s42984-020-00001-2).

Palombini, C. (1993) Pierre Schaeffer, 1953: towards an experimental music. *Music & Letters*. 74 (4), 542–557.

Parali, F. (n.d.) *Fani Parali*. Fani Parali. <https://faniparali.org/performance> [Accessed: 13 August 2022].

Parikka, J. (2015) *The Anthrobscene*. Minneapolis, MN: University of Minnesota Press.

Parker, J.E.K. & Dockray, S. (2023) ‘All possible sounds’: speech, music, and the emergence of machine listening. *Sound Studies*. 9 (2), 253–281. doi:[10.1080/20551940.2023.2195057](https://doi.org/10.1080/20551940.2023.2195057).

Pasquinelli, M. (2021) How to make a class: Hayek’s neoliberalism and the origins of connectionism. *Qui Parle: Critical Humanities and Social Sciences*. 30 (1), 159–184.

Peng, K., Mathur, A. & Narayanan, A. (2021) Mitigating dataset harms requires stewardship: lessons from 1000 Papers. *arXiv:2108.02922 [cs]*. <http://arxiv.org/abs/2108.02922>.

Pestana, N. (n.d.) *Nestor Pestana*. Nestor Pestana. <https://nestorpestana.com> [Accessed: 13 August 2022].

Phan, T. (2022) *Machine Listening: Thao Phan: Listening to Misrecognition*. [YouTube video]. <https://www.youtube.com/watch?v=D3vbd4QHeb0>.

Phillips, R. & Abbas-Nazari, A. (2022) Fostering natural world engagements: design lessons and issues from the My Naturewatch Training Program. *Design Issues*. 38 (2), 47–63. doi:[10.1162/desi_a_00681](https://doi.org/10.1162/desi_a_00681).

Pohflepp, S. (n.d.) *Passings of the Flame*. https://www.academia.edu/19758339/Passings_of_the_Flame.

Ponterotto, J. (2006) Brief note on the origins, evolution, and meaning of the qualitative research concept ‘thick description’. *Qualitative Report*. 11, 538–549. doi:[10.46743/2160-3715/2006.1666](https://doi.org/10.46743/2160-3715/2006.1666).

Potter, C. (2020) West African voice-disguisers and audible ghosts: a case for expanding the fluency of global design history. *Design and Culture*. 12 (3), 309–329. doi:[10.1080/17547075.2020.1741912](https://doi.org/10.1080/17547075.2020.1741912).

Prado, L. & Oliveira, P. (2015) *Futuristic gizmos, conservative ideals*. Modes of Criticism. [Online]. 27 February. Available from: <https://modesofcriticism.org/futuristic-gizmos-conservative-ideals/> [Accessed: 17 June 2021].

Preece, J., Sharp, H. & Rogers, Y. (2015) *Interaction design: beyond human-computer interaction*. 4th edition. Chichester: John Wiley.

Piringer, J. (2019) *darkvoice*. Bandcamp. <https://joergpiringer.bandcamp.com/album/darkvoice>.

Pugliese, J. (2010) *Biometrics: bodies, technologies, biopolitics*. New York: Routledge.

Pullin, G. & Hennig, S. (2015) 17 ways to say yes: toward nuanced tone of voice in AAC and speech technology. *Augmentative and Alternative Communication*. [Online] 31 (2), 170–180. Available from: doi [10.3109/07434618.2015.1037930](https://doi.org/10.3109/07434618.2015.1037930).

QuikQuak (n.d.) *Audio plug-ins for PC & Mac. FX and Synths*. QuikQuak. <https://www.quikquak.com/> [Accessed: 27 November 2022].

Radovic, S. (2008) *First computer to sing - Daisy Bell*. [YouTube video] <https://www.youtube.com/watch?v=41U78QP8nBk>.

Rangarajan, S. (2021) Hey Siri—why don’t you understand more people who talk like me?. *Mother Jones*. February 23. <https://www.motherjones.com/media/2021/02/digital->

[assistants-accents-english-race-google-siri-alexa/](#) [Accessed: 15 August 2023].

Reed, M.S. (2018) *The research impact handbook*. 2nd edition. S.l.: Fast Track Impact.

Replica Studios (n.d.) *Synthesize voice AI and natural sounding text-to-speech – Replica*. [Online]. Replica Studios. Available from: <https://replicastudios.com/> [Accessed: 18 August 2021].

Ridler, A. (2018) *Myriad (Tulips)*. Anna Ridler. <http://annaridler.com/myriad-tulips> [Accessed: 5 May 2022].

Roemmele, B. (2016) *The Voder: 1939, the world's first electronic voice synthesizer*. [YouTube video]. https://www.youtube.com/watch?v=TsdOej_nC1M. [Accessed: 5 May 2022]

Rogers, C.R. & Farson, R.E. (2021) *Active listening*. Bristol: Mockingbird Press.

Romaine, S. (2017) Language endangerment and language death. In A. F. Fill & H. Penz (eds.). *The Routledge handbook of ecolinguistics*. New York: Routledge, pp. 40-55.

Rosner, D. (2018) *Critical fabulations: reworking the methods and margins of design*. Cambridge, MA: MIT Press.

Sanders, E.B.-N. & Stappers, P.J. (2008) Co-creation and the new landscapes of design. *CoDesign*. 4 (1), 5–18. doi:[10.1080/15710880701875068](https://doi.org/10.1080/15710880701875068).

Schaeffer, P. (2017) *Treatise on musical objects: an essay across disciplines*. Translated by C. North and J. Dack. Berkeley, CA: University of California Press.

Schafer, R., M. (1977). *The tuning of the world*. New York: A.A. Knopf.

Schlichter, A. & Eidsheim, N.S. (2014) Introduction: voice matters. *Postmodern Culture: Journal Of Interdisciplinary Thought On Contemporary Cultures* | [Online] 24 (3). Available from: <http://www.pomoculture.org/2017/09/09/introduction-voice-matters/> [Accessed: 5 September 2021].

Schmid, H. (2017) *Uchronia: time at the intersection of design, chronosociology and chronobiology*. PhD thesis, Royal College of Art. <https://researchonline.rca.ac.uk/2748/>.

Schön, D.A. (2016) *The reflective practitioner: how professionals think in action*. New York, NY: Routledge

Science Museum (n.d.) *Goodbye to the hello girls: automating the telephone exchange*. Science Museum. <https://www.sciencemuseum.org.uk/objects-and-stories/goodbye-hello-girls-automating-telephone-exchange> [Accessed: 19 August 2022].

Semel, B. (2020) *The body audible: from vocal biomarkers to a phrenology of the throat*. Somatosphere. [Online]. Available from: <http://somatosphere.net/2020/the-body-audible.html/> [Accessed: 4 October 2021].

Semel, B. (2021) *A new AI lexicon: voice*. [Online]. AI Now, 28 October. Available from: <https://ainowinstitute.org/publication/a-new-ai-lexicon-voice> [Accessed: 29 October 2021].

Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M. & Villalobos, P. (2022) Compute trends across three eras of machine learning. *arXiv*. 2202.05924. doi:[10.48550/arXiv.2202.05924](https://doi.org/10.48550/arXiv.2202.05924).

Sinders, C. (n.d.) *Feminist data set*. Caroline Sindere. <https://carolinesinders.com/feminist-data-set/> [Accessed: 5 May 2022].

Singh, R. (2012) *Rita Singh*. Machine Learning For Signal Processing Group, Carnegie Mellon University, 2 November. <http://mlsp.cs.cmu.edu/people/rsingh/index.html> [Accessed: 31 August 2022].

Singh, R. (2019) *Profiling humans from their voice*. Dordrecht: Springer.

Sorry to Bother You (2018) Directed by Boots Riley. USA: Annapurna Pictures.

Sorry to Bother You (2018) Directed by Boots Riley. USA: Annapurna Pictures. (Clip) [YouTube video] <https://www.youtube.com/watch?v=T5X3cu1B87k>.

Soundtoys (n.d.) *Crystallizer*. Soundtoys. <https://www.soundtoys.com/product/crystallizer/> [Accessed: 12 September 2023].

Squeaky Wheel (2021) *Johann Diedrick: Dark Matters*. Squeaky Wheel Film & Media Art Center. <https://squeaky.org/event/johann-diedrick-dark-matters/>.

Sterne, J. (2003) *The audible past: cultural origins of sound reproduction*. Durham, NC: Duke University Press.

Sterne, J. (ed.) (2012) Sonic imaginations. In *The Sound Studies Reader*. New York: Routledge.

Sterne, J. (2015) Space within space: artificial reverb and the detachable echo. *Grey Room*. (60), 110–131. doi:[10.1162/GREY_a_00177](https://doi.org/10.1162/GREY_a_00177).

Sterne, J. & Sawhney, M. (2022) The acousmatic question and the will to datafy: Otter.ai, low-resource languages, and the politics of machine listening. *Kalfou*. 9 (2). <https://tupjournals.temple.edu/index.php/kalfou/article/view/617>.

Stoeve, J.L. (2016) *The sonic color line: race and the cultural politics of listening*. New York:

NYU Press.

Superflux (2017) *Our Friends Electric*. Superflux. <https://superflux.in/index.php/work/friends-electric/> [Accessed: 2 January 2024].

Supperware (n.d.) *Supperware*. <https://supperware.co.uk> [Accessed: 24 November 2023].

Sutton, S.J., Foulkes, P., Kirk, D. & Lawson, S. (2019) Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. [Online]. 2 May 2019 New York, NY, USA, Association for Computing Machinery. pp. 1–14. Available from: doi:[10.1145/3290605.3300833](https://doi.org/10.1145/3290605.3300833) [Accessed: 25 March 2021].

Tate (2019) *Summer School 2019*, Tate Modern. Tate. <https://www.tate.org.uk/whats-on/tate-modern/summer-school-2019> [Accessed: 12 April 2023].

Tate, L. (n.d.) *The difference between speech and voice recognition*. Kardome. <https://kardome.com/blog-posts/difference-speech-and-voice-recognition> [Accessed: 10 November 2022].

Taylor, A. (2018) The automation charade. *Logic Magazine*, August 1. [Online]. Available from: <https://logicmag.io/failure/the-automation-charade/> [Accessed: 7 October 2020].

TECHNE (n.d.) *NPIF Students - Techne AHRC Doctoral Training Partnership*. TECHNE <http://www.techne.ac.uk/for-students/techne-students/techne-npif-students> [Accessed: 15 August 2022].

Terralingua (n.d.) *What is biocultural diversity?* Terralingua. <https://terralingua.org/what-we-do/what-is-biocultural-diversity/>.

Thomson, P. (2019) Summer school day three. *patter*. Pat Thomson. <https://patthomson.net/2019/08/01/summer-school-day-three-2/>.

Tompkins, D. (2011) *How to wreck a nice beach: the vocoder from World War II to hip-hop*. New York: Melville House Publishing.

TransVoiceLessons (2021) *Voice Feminization for ABSOLUTE BEGINNERS | How to Get Started Now*. [YouTube video] <https://www.youtube.com/watch?v=BfCS01MkbIY>.

TTP (2021) *Amazon's data dragnet*. Tech Transparency Project. <https://www.techtransparencyproject.org/articles/amazons-data-dragnet> [Accessed: 5 September 2023].

Twenge, J.M. (2023) How Gen Z changed its views on gender. *Time*, 1 May. <https://time.com/6275663/generation-z-gender-identity/> [Accessed: 24 November 2023].

Unit Test (2023) *Not I*. Creative Informatics. <https://creativeinformatics.org/research/not-i/>.

Vallee, M. (2017) Technology, embodiment, and affect in voice sciences: the voice is an imaginary organ. *Body & Society*. 23 (2), 83–105. doi:[10.1177/1357034X17697366](https://doi.org/10.1177/1357034X17697366).

Van Leeuwen, T. (2016) A social semiotic theory of synesthesia? A discussion paper. *HERMES - Journal of Language and Communication in Business*. 105. doi:10.7146/hjlc.v0i55.24292.

Vieira de Oliveira, P.J.S. (2021) “...the table was set, and we were never dead”: On the persistence of colonial listening in Germany. *MAST*. 2 (2), 89–101.

Vocalis Health (n.d.) *VocalisCheck – Vocalis Health*. [Online]. Available from: <https://vocalishealth.com/vocalis-health-products/vocalischeck> [Accessed: 17 September 2021].

Voegelin, S. (2014) *Sonic possible worlds: hearing the continuum of sound*. New York: Bloomsbury.

Voegelin, S. (2019). Max Reinhardt with Salome Voegelin. *Late Junction*, BBC Radio 3. [Online]. Available from: <https://www.bbc.co.uk/sounds/play/m0003c70> [Accessed: 15th April 2019].

Warren, T. (2020) *Zoom faces a privacy and security backlash as it surges in popularity*. The Verge, 1 April. <https://www.theverge.com/2020/4/1/21202584/zoom-security-privacy-issues-video-conferencing-software-coronavirus-demand-response> [Accessed: 12 May 2022].

Watson, S. (2019) The unheard female voice. [Online]. 15 12:28:30 2019. *The ASHA Leader*, 1 February. Available from: doi:[10.1044/leader.FTR1.24022019.44](https://doi.org/10.1044/leader.FTR1.24022019.44) [Accessed: 31 October 2021].

Walshe, J. (2019) *Ghosts of the hidden layer*. Milker Corporation. [Online]. Available from: <http://milker.org/ghosts-of-the-hidden-layer> [Accessed: 22 April 2021].

Weitzman, C. (2022) *Deepfake voice*. Speechify. <https://speechify.com/blog/deepfake-voice/> [Accessed: 5 January 2024].

Wen, Y., Raj, B. & Singh, R. (2019) Face reconstruction from voice using generative adversarial networks. In: H. Wallach, et al. (eds.). *Advances in neural information processing systems*. [Online]. Available from: <https://proceedings.neurips.cc/paper/2019/file/eb9fc349601c69352c859c1faa287874-Paper.pdf>.

West, M., Kraut, R. & Chew, H.E. (2019) *I'd blush if I could: closing gender divides in digital skills through education*. [Online]. EQUALS; UNESCO; Available from: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1> [Accessed: 11 December 2020].

Wigmore, J. (1926) A new mode of identifying criminals. *Journal of Criminal Law and Criminology*. 17 (2), 165.

Wikipedia. (2021) *International Prototype of the Kilogram*. Wikipedia. [Online]. Available from: https://en.wikipedia.org/wiki/International_Prototype_of_the_Kilogram [Accessed: 28 October 2021].

Woolard, K. (2004) 'Codeswitching'. In A. Duranti (eds.) *A companion to linguistic anthropology*. Malden and Oxford: Blackwell, pp. 73– 94.

Yollin, P. (2007) Elephants can hear through their feet / Researcher says ability more useful to animals in wild. *SFGATE*, 19 May. <https://www.sfgate.com/science/article/elephants-can-hear-through-their-feet-2593114.php> [Accessed: 24 April 2023].

Zoom Support (2021) *Creating an interactive voice response system*. 2021. Zoom Support. <https://support.zoom.us/hc/en-us/articles/360038601971-Creating-an-interactive-voice-response-system> [Accessed: 24 April 2023].

Zuboff, S. (2019) *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. New York: PublicAffairs.