Communing with the Black Box:
Using publicly accessible AI audio tools to generate rich sonic outcomes in the composition process.

Philly Holmes

Submitted as part of the MSc in Sound Design, Edinburgh College of Art

August 2022

THE UNIVERSITY of EDINBURGH
Edinburgh College of Art

Made possible by funding from the Andrew Grant Scholarship & Bequest

Image courtesy of OpenAI Dall-E 2 (August 2022)

THE UNIVERSITY *of* EDINBURGH
Edinburgh College of Art

Communing with the Black Box:
Using publicly accessible AI audio tools to generate rich sonic outcomes in the composition process.

Philly Holmes

Submitted as part of the MSc in Sound Design, Edinburgh College of Art

August 2022

Made possible by funding from the Andrew Grant Scholarship & Bequest.

Word Count: 7469

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

Signed: Philip Holmes

Abstract & Lay Summary

In the past decade, deep learning/machine learning/artificial intelligence techniques have emerged in the audio industry, producing a broad spectrum of sonic outcomes, and unlocking potential for complex abstract sound processing techniques. These tools span from complex audio source separation to straightforward MIDI generation techniques. In actuality, modern deep learning 'Artificial Intelligence' (AI) tools are just a new technological entry into a lineage of algorithmic techniques for musical composition; algorithms and formal rulesets have been a crucial part of the compositional process for centuries.

The development of modern computing technology takes algorithmic composition to the extreme and abstract. The application of complex, highly abstracted processes to musical data results in unique, often unpredictable, results that exist outside of the scope of straightforward human capability or imagination – the neural network black box.

This project takes a practice-led approach to these tools and engages with deep learning as an extension of creative workflows. seeking to explore the nature of existing AI audio tools, with a focus on plug-and-play, easily available audio processing techniques. Through a set of experiments and case studies, various tools are assessed for their usefulness in creative workflows. These are presented as audio pieces, performance recordings, compositions and work created in Unity. Key projects in the field are also highlighted and presented, showcasing important projects at the vanguard of AI audio.

Acknowledgements

I'd like to, firstly, thank my family for continuous support during my masters studies. Without their encouragement I would not be where I am today. Thank you mum, your world class skills as a proofreader and outside reader have shaped my academic work and taught me to communicate my ideas effectively. Thank you dad, who taught me to approach technology with playfulness always and gave me my love for strange music.

To all of the staff at the Reid School of Music, I cannot thank you enough. To Dr. Martin Parker whose undiluted enthusiasm for the sonic arts continues to inspire hundreds of students every year, empowering each and every one of us to become better artists and creatives. The opportunities he provided me this year have fundamentally reshaped my practice for the better and for that I am endlessly grateful.

To Roderick Buchanan-Dunlop and Louis McHugh, a tour de force team of audio technology wizards who have provided endless technical support and assistance throughout the year. Their skills help to enable the technical execution of so much incredible work. They're the backbone of every technical project in the Reid School of Music.

To Dr. Chris Letcher and Dr. Tom Mudd, who took time to read my drafts and send wonderfully detailed feedback, helping to push the quality of this project to the highest level. Thank you.

And to Dr. Jules Rawlinson, without whom this project would not exist. He was the first person to meet with me about the initial ideas for the precursor to this project. Without a tangential discussion about machine learning tools for audio processing, this project simply would not exist in its current form. His boundary pushing work is endlessly inspirational and his enthusiasm for experimental work is infectious. Thank you.

Thank you to the Andrew Grant Scholarship and Bequest, an award I do not take lightly, for funding my studies this year and giving me the freedom to experiment.

To Dearbhla, my best friend and absolute rock. Your endless support for my silly ideas has kept me afloat all year. Real.

To tireless crew of hospitality staff in Cowgate and in Paradise Palms. Thank you, we're in this together.

And finally, endless thanks to Nani, Elliot, Bryant, Alliyah, Sophie, Mark, Megan and Eoin, the PlusMinus Ensemble, Oisín, Anna and the Femmergy family, the Miss World crew, Roo, Aoife and the Club Comfort crew, Jamie, Feena and everyone at EHFM, and Rowan and the whole team at Palms Records.

To the moon and stars and back forever.

Table of Contents

List of Figures

Note on Supplemental Material

An extensive database of Supplemental material is included alongside this text in the form of appendices. A list of appendices can be found in the back matter of this document.

This material is presented throughout the text, and it's therefore necesssary to surface this media at the beginning of this document. It is recommended that examples are listened to as they are mentioned in the text. Appendices are listed in order of appearance.

You can find the collected material at the following google drive link (August 2022): https://drive.google.com/drive/folders/10399wLzS66PLd2iMgxjwQTUBBbuZOgKj?usp=sharing

Or at the following OneDrive Link (August 2022): https://uoe-my.sharepoint.com/:u:/g/personal/s2228269_ed_ac_uk/EfIkk3QxacpOshvgenPZx3QBpmYqKSycWWX8IAq8dAIvXQ?e=tfrjHG

If this link does not work, email pholmes@tcd.ie or s2228269@ed.ac.uk to request access.

Communing with the Black Box:
Using publicly accessible AI audio tools to generate rich sonic outcomes in the composition process.

Introduction

In the past decade, deep-learning techniques in audio have emerged as a novel frontier for sonic exploration, producing a broad spectrum of outcomes and unlocking potential for complex voice-recognition and re-synthesis techniques, audio-source separation and any number of predictive MIDI-generation tools. In actuality, modern deep-learning 'Artificial Intelligence' (AI) tools are just a new technological entry into a lineage of algorithmic techniques for musical composition; algorithms and formal rulesets have been a crucial part of the compositional process for centuries.

Modern computing power and machine learning take algorithmic composition to the extreme, applying complex, abstracted processes to musical data, creating unique,[1] often unpredictable, results that exist outside of the scope of straightforward human capability or imagination – the neural network black box.

This project aims to explore current, publicly available AI audio tools from a practitioner's perspective, using these tools as part of the creative process, embracing their quirks and idiosyncrasies and allowing them to emerge with their own unique sonic imprint. These explorations will be presented in the form of sonic sketches, samples, compositions and semi-interactive game-engine musical work created using Unity.

The project's inspiration originates from the rising popularity of AI technology in public discourse and a desire to explore AI tools in a sonic context, intersecting digital compositional practices with novel machine-learning techniques and asking the question 'How can we use AI audio tools in the creative process?'

Developments in neural-network image generation have made leaps and bounds in the past decade – Google's *DeepDream* from 2015[2] looks primitive compared to tools such as *Dall-E 2*, which is currently in a beta form but creates incredibly detailed imagery based on language-parsing text prompts and now functions as a set of fully fledged creative tools.[3] In comparison, AI audio tools are still nascent, less accessible and less capable of producing instantly workable results. This is likely an issue of fidelity – AI audio has a significantly higher resource cost, as each second of audio is composed of many thousands of samples. As Spratley et al posit:

---

1 Nikita Braguinski, *Mathematical Music: From Antiquity to Music AI / Nikita Braguinski.* (Oxford, England ; Routledge, 2022), 76.

2 Alexander Mordvintsev, Christopher Olah, and Mike Tyka, 'DeepDream - a Code Example for Visualizing Neural Networks', *Google AI Blog* (blog), 1 July 2015, http://ai.googleblog.com/2015/07/deepdream-code-example-for-visualizing.html.

3 Aditya Ramesh et al., 'Hierarchical Text-Conditional Image Generation with CLIP Latents' (arXiv, 12 April 2022), https://doi.org/10.48550/arXiv.2204.06125.

> ‘Capturing high-level image features might require deeper neurons to be receptive to say 200 pixels squared. A high-level audio feature may be sustained over a second increasing the receptive fidelity requirement to tens of thousands of time stamps.’[4]

The production of AI audio requires high-fidelity, sustained over an extended time domain, a data-intensive process. As neural network and computing processes become more streamlined, it’s likely that we’ll see major improvements in AI audio fidelity and process over the next few years. The tools used in this project represent the start of emerging improvements in AI audio.

Like any endeavour in audio, using these tools is labour-intensive, requiring a huge amount of creative and curatorial intervention to produce compelling results. As sonic experimentalist Holly Herndon states, ‘AI is just us. AI is human labor obfuscated through a terminology called AI’.[5] Current forms of these tools lack the capability of contextualisation that is required for musical meaning-making and cannot yet be a stand-in or replacement for creativity or creation; they exist as another digital musical process among thousands.[6]

Discussion of artificial intelligence is inextricably linked to discussions of ethics, data usage, copyright and labour.[7] Thorough exploration of these subjects falls outside of the scope of this essay but will arise as an inevitable part of the process of engaging with large dataset-based machine-learning tools and will be addressed where possible.

---

4 Steven Spratley, Daniel Beck, and Trevor Cohn, ‘A Unified Neural Architecture for Instrumental Audio Tasks’ (arXiv, 28 February 2019), https://doi.org/10.48550/arXiv.1903.00142.

5 Emily McDermott, ‘Holly Herndon on Her AI Baby, Reanimating Tupac, and Extracting Voices’, *ARTnews.Com* (blog), 7 January 2020, https://www.artnews.com/art-in-america/interviews/holly-herndon-emily-mcdermott-spawn-ai-1202674301/.

6 Shelly Knotts and Nick Collins, ‘A Survey on the Uptake of Music AI Software’, *Proceedings of the International Conference on New Interfaces for Musical Expression* (Birmingham, UK, Zenodo, 1 June 2020), 502, https://doi.org/10.5281/zenodo.4813499.

7 Sara Brown, ‘The Hidden Work Created by Artificial Intelligence Programs’, MIT Sloan, 4 May 2022, https://mitsloan.mit.edu/ideas-made-to-matter/hidden-work-created-artificial-intelligence-programs.

Chapter One: Finding a Voice

Identifying AI audio tools and evaluating their usefulness in a creative process.

Machine-learning music tools and processes designed to generate and manipulate sonic material are increasingly abundant and often freely available. In the commercial realm, companies such as *iZotope*, responsible for industry standard audio-repair and post-processing tools, are implementing machine learning-based processes to their software to improve production workflows. These tools deploy AI-informed audio analysis processes to guide and optimise effects parameters.[8] For example, *Amper Music* and *AIVA* claim to be replacing the composer, creating 'original' neural network-generated material at the click of a button. The adoption of machine-learning techniques into preexisting audio production workflows is continuing to grow as companies and creators alike realise the potential labour-saving potential for these products.[9] Extensive exploration of these kinds of labour-saving AI audio tools falls outside the scope of this project, as they typically do not produce any transformative or unique sonic outcomes but instead focus on producing human-replicable outcomes on a larger scale or replacing the composer entirely. They serve to supersede creative sonic workflows rather than assisting with them.

In the open-source realm, machine learning for music is saturated with tools that rely on data-pattern recognition and MIDI/musical notation. These tools use the predictability and fixed nature of digital notation combined with high-fidelity virtual or sample-based instruments to produce impressive-sounding results. These results are easy to replicate by hand in any digital audio workstation. Tools such as Google Magenta's *Bach Google Doodle* [10] or *OpenAI's MuseNet* [11] create soundalikes derived from the analysis of existing datasets of popular artists. The ability to generate Mozart's 'Alla Turca' in the style of Lady Gaga is novel, [12] but it fundamentally does not unlock new ways to create audio or offer something legally sampleable for the artist engaging with these tools. As Sollit notes, 'Magenta has been able to generate musical compositions, but

8 David Bawiec, 'IZotope and Machine Learning: Speeding Up Your Workflow with Assistive Audio Technology', iZotope, 9 January 2020, https://www.izotope.com/en/learn/speed-up-your-workflow-with-assistive-audio-technology.html.

9 Knotts and Collins, 'A Survey on the Uptake of Music AI Software', 500.

10 Cheng-Zhi Anna Huang et al., 'Coconet: The ML Model behind Today's Bach Doodle', Blog, Magenta, 20 March 2019, https://magenta.tensorflow.org/coconet.

11 Christine McLeavey Payne et al., 'MuseNet', OpenAI, 25 April 2019, https://openai.com/blog/musenet/.

12 Ibid.

lacks the innate story that comes from humans when they interpret and perform the score.'[13]

Most publicly available machine-learning music tools exist in this form, with the goal of creating sonic doppelgangers of existing music. While novel and often uncanny in output, as creative tools they are impractical. They offer very little in terms of compelling or practical sonic output. A cynical reading of the field could suggest that many of these soundalike tools exist as marketing tools to showcase the power of each company's AI capabilities, as none of them investigate how their technologies can be applied to creative musicking. A milder reading could suggest that we require these attention-grabbing, but still-primitive tools to convince users of the value of the technology, producing user buy-in for more creative and expressive tools down the road.[14] Furthermore, these soundalike tools prompt questions of copyright. Who gets to claim ownership of the newly generated material output of an algorithm when it was trained on preexisting material in the private domain? The question of ownership makes it almost impossible to engage with these specific soundalike tools as a creative practitioner, as to do so could risk the ire of copyright and intellectual property law.

To avoid these pitfalls, four loose criteria were established to identify practical tools with usable results for this project.

First, the AI tools had to take direct user input of some form. As described above, many of these tools are only capable of limited user input,[15] often only preexisting categories of training data. In their current form, such tools are not conducive to creating original material. Inversely, the few tools that take user audio or text input, described later in this essay, provide limitless and varied original output reflective of or responsive to the source input and therefore compelling and usable.

Second, these tools needed to output audio. Direct audio processing exposes creative possibilities that other techniques are simply not capable of. Predictive MIDI tools that output notation of various levels of structural coherence could be reproduced by any number of algorithmic or analytical compositional techniques. They're a form of digital serialism.[16] On the other hand, machine-learning tools that output audio are capable of numerous abstract and complex processes that are otherwise impossible to execute.

Third, the tools had to be publicly accessible and easy to use. Many AI audio techniques

13 Lucy Sollit, 'Collaborating with Intelligent Machines', *Intersections: Art and Digital Creativity in the UK* (blog), 27 April 2017, https://medium.com/intersections-arts-and-digital-culture-in-the-uk/collaborating-with-intelligent-machines-cb5ecf32c98d.

14 Sollit; Braguinski, *Mathematical Music*, 82.

15 Huang et al., 'Coconet'.

16 Braguinski, *Mathematical Music*, 78.

require a high investment of time, data and labour. Whereas huge technology corporations have access to the resources required to amass generate training data, creative practitioners rarely do. For example, Deezer's *Spleeter*, which will be explored in more detail in this project, functions better than other algorithms with a similar purpose because of Deezer's access to an extensive library of licensed and copyrighted music.[17] For the purposes of this project, tools with a relatively low barrier to entry were chosen. As with any creative tool or process, time investment is required to experiment and find best practices. Consequently, to maximise this experimentation, tools that produced output with a low barrier to entry were prioritised.

Fourth, the results had to be compelling, or noticeably unique. The word 'compelling' is used throughout this essay as a means to describe sounds that can't necessarily be described as 'good' or harmonically pleasing in a traditional sense, but carry a (subjective) engaging quality, interesting timbral quality or unique energy that can't necessarily be otherwise categorised – a sonic quality that draws you in. A selection of currently available AI audio tools are designed to sound indistinguishable from real, recorded samples.[18] They rarely produce differentiable or unique results and don't appear to offer anything more than a different workflow to explore common sounds. This project aims to explore the unique fringes and limits of the technology, and for this reason, tools that produced results indistinguishable from non-deep learning sampling techniques, while valuable, were excluded.

These four categories offer a framework for engaging with AI audio tools that prioritises user input and ease of use with low technical resource requirements. This ensures that they can be integrated into the creative process with ease while setting appropriate scope for this project. The rest of this chapter will describe several tools that meet these criteria and explore their value as novel sound design processes.

*Spleeter*___

Spleeter is a music source separation library developed by Deezer for the purposes of retrieving or extracting separate tracks or stems from a multitrack mix.[19] It uses *TensorFlow*, an open-source Python-based machine-learning library developed by Google.[20] *Spleeter* was

17 Manuel Moussallam, 'Releasing Spleeter: Deezer R&D Source Separation Engine', Medium, 3 February 2020, https://deezer.io/releasing-spleeter-deezer-r-d-source-separation-engine-2b88985e797e.

18 Lyubomir Dobrev, 'AudiaLab Synapse Drums Creates Drum Samples with Artificial Intelligence', gearnews.com, 10 February 2022, https://www.gearnews.com/audialab-synapse-drums-creates-drum-samples-with-artificial-intelligence/.

19 Moussallam, 'Releasing Spleeter'.

20 Martin Abadi et al., 'TensorFlow: A System for Large-Scale Machine Learning', in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, 265–83, https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.

trained using Deezer's extensive streaming catalogue but permits users to train their own models using the same framework with an alternative dataset.[21] For the purposes of this project, the original Deezer models were used. The resources required to develop a dataset capable of training this algorithm to a usable standard are outside of the scope of this project.

Spleeter relies on time-frequency masking techniques[22] to extract stems. The resulting files sound like other manual process that deploy this technique. Most similar is the so-called 'Youtube Acapella', a DIY acapella created from a mastered mp3 that shows clear evidence of attempts to filter and mask the instrumental parts of a track.[23] The resulting sounds have an 'underwater' quality, with imperfect artefacts of the instrumental bubbling in the background. *Spleeter* produces similar- sounding results at a more granular scale, outputting up to 5 audio tracks across different categories of sound. *Spleeter* and other source separation tools have found use in the DIY remix communities of *SoundCloud*, where artists use online implementations of *Spleeter* to retrieve vocal and instrumental sections of songs for the purposes creating their own remixes and unofficial edits.

As a tool designed for sonic source separation, it was unclear whether *Spleeter* could produce unique or compelling sonic results that extended past its intended purpose. As a test case for the creative viability of this tool, a set of 10 short, gestural tracks and sonic explorations were composed. These were processed using *Spleeter*'s 5-stem output and then combed for samples and sonic material of various lengths and timbral qualities. It quickly became clear that *Spleeter*'s algorithm was capable of surfacing unique samples based on original material, revealing atypical combinations of sound based on the source material. At this point, *Spleeter*'s 'voice' emerged – the previously described 'underwater', comb filtering quality. This degraded sound produced myriad rich and compelling sonic outcomes, ideal for resampling. The sonic explorations, resulting stems and curated samples are all contained in the files for *Appendix A*.

The compelling outcomes from the initial sketches proved a valid test case to expand the corpus and create a larger set of samples and snippets. Approximately 150 unreleased demos, tracks and experiments dating from 2018 to 2022 were exported, ordered and batch-processed using *Spleeter*, with the intention to create a large corpus of samples from which to work with in a compositional context. Each of the stems resulting from the batch process had to be manually auditioned for samples. Approximately 970 samples and snippets were derived from 750 stems. In raw form, these stems are too cumbersome to repeatedly audition in full in a creative context. Machine-learning tools such as *Spleeter* are incapable of creatively curating sound in a practical

21 Moussallam, 'Releasing Spleeter'.
22 Moussallam.
23 'Vocal Removal Plug-Ins - Audacity Wiki', Wikipedia, Audacity Team Wikipedia, 13 October 2020, https://wiki.audacityteam.org/wiki/Vocal_Removal_Plug-ins.

way, and using a large-scale, manual-sampling process was the only clear way to derive usable results from over 20 hours of source-separated audio. Despite the labour-intensive nature of this process, the resulting sample library contains a vast array of compelling timbres ready to be re-composed into new material and offers a very accessible means to engage with machine-learning tools in the compositional process. This sample corpus is contained in *Appendix B*, labelled according to track number, stem category from which the sample came and subjective timbral descriptor. The source tracks and resulting stems have been omitted from submission due to file size restrictions. A quick scan through this corpus' files would be the best means of navigation in the presented appendix. It would best allow someone to understand the timbral variety at play in this library. The analytical tools describe below may also be useful for navigating such a large body of samples.

*Magenta DDSP*___

Magenta DDSP, Google Magenta's 'live tone transfer tool', takes live audio and applies 'tone transfer' processes to transform source audio into another instrument or sonic fingerprint. For example, a sung melody can be processed and imbued with the qualities a clarinet, tuba or violin.[24] Currently, the tool is in the early stage of development and produces primitive-sounding results, but 'tone transfer' technology unlocks a vast array of musical potential. At first glance, it could potentially enable non-instrumentalists to create music using one device and have it sound like another. It also opens other avenues of exploration. For example, the musical muscle memory of a guitarist is inherently different to that of a vocalist – the interfaces used to produce sound privilege certain techniques and musical shapes over others. 'Tone transfer' technology could allow for a vocalist to take their musical decision-making and morph it to sound like a guitar, creating a hybrid sonic fingerprint with uncanny qualities of both vocals and guitar. This technology also has experimental application. The newest *Magenta DDSP* audio plugin takes live source audio and algorithmically processes it, outputting a 'monophonic' tone transfer. By pushing this technology to its limit and using polyphonic and complex sound textures, a musician can create chaotic and incredibly unpredictable sonic clusters as the algorithm grapples with pitch detection.

The 'Tone Transfer Experiments' contained in *Appendix C* demonstrate the timbral variety of *Magenta's* tools. A monophonic and polyphonic musical snippet played using both a simple sine wave and complex piano sample instrument were used to showcase the transformative effect of

24 Jesse Engel et al., 'DDSP: Differentiable Digital Signal Processing', Magenta, 15 January 2020, https://magenta.tensorflow.org/ddsp.

these tools. The results from processing the monophonic ‘Source A’ file illustrate the capability of tone-transfer technology. Despite morphing a clean tone with a variety of instrumental profiles, the resulting sound has a raspy, saxophone-like sound quality. This persists throughout the prepopulated instruments in this tool, with even the vocal and violin profiles retaining a very present woodwind quality. The processed ‘Source C’ piano files demonstrate how the technology struggles with more complex timbral inputs. The tool struggles to separate the fundamental frequency from the piano harmonics, and the resulting sounds retain a woodwind quality with a significantly more garbled and inconsistent digital tone. While it’s clear the tool breaks down with more complex sound inputs, the results remain interesting. The garbled, sliding tones of the processed ‘Source C’ are uncanny and, despite illustrating the limits of this still-nascent technology, produce compelling and certainly usable results. Tonal transfer tools are improving at a rapid pace, and while it’s unlikely that these toolsthey will be able to produce results indiscernible from real performers in the immediate future, they present a new avenue for creative audio creation powered by machine learning.

Holly Plus___

Holly Plus is a non-real-time ‘vocal tonal transfer tool’ modelled on the voice of sonic artist and performer Holly Herndon. Developed with Herndon by Never Before Heard Sounds, it uses similar tonal transfer technology to the Google *Magenta DDSP* tools mentioned above.

> ‘A Voice Model is a deep neural network that can generate raw audio of an individual voice. The network is trained on recorded speech and singing from the target voice, and can be interacted with in various ways, from text-to-speech applications to more complex interactions such as audio style transfer, where audio from one voice can be converted to resemble the target voice, a kind of vocal puppetry.’[25]

Put simply, it takes audio files uploaded online and processes them using Herndon’s digital vocal fingerprint. The result is an uncanny and alien but still distinctly familiar sound with a uniquely AI-sounding quality. *Holly Plus* produces engaging results almost instantly and has the lowest barrier to entry of any tool discussed in this essay. Herndon’s work on this tool serves as a clear test case for the use of AI music tools for new, transformative sonic exploration. The experiments in *Appendix C*’s experiments contain examples of simple audio processed using *Holly Plus*. When compared to *Magenta DDSP*’s monophonic outputs, Herndon’s *Holly Plus* handles both simple and relatively complex audio impressively. Both ‘Source A’ and ‘Source B’ remain clear when processed and don’t contain any of the pitch-quantization artefacts found

25 Holly Herndon, ‘Holly+ 👭 🗣 💪’, 13 July 2021, https://holly.mirror.xyz/54ds2IiOnvthjGFkokFCoaI4Ea-bytH9xjAYy1irHy94.

in the Magenta outputs. Despite ‘Source C’ and ‘Source D’s relative complexity, they are also handled well, producing noticeably different but still recognisable results. *Appendix D* contains snippets of my speaking voice processed using *Holly Plus*. It was interesting to see how the tool breaks down when processing relatively dry, nonmusical narration, struggling to discern a root note from speech. It fails to retain any words or discernible meaning in its processing. It is important to note that this tool is still in development, and a recent public showcase of a new version of *Holly Plus* at Sónar Festival 2022 demonstrates the tool processing vocals in real-time at a much higher fidelity than the current online tool.[26] *Holly Plus* is a versatile tool at the vanguard of machine-learning audio.

Processing and Analysis___

Using machine-learning tools, attempts were made to analyse the corpus of samples produced with *Spleeter*. The timbral variance of each sample provided a unique opportunity to explore a variety of different analytical tools. On reflection, these tools did not provide any statistical conclusions about the corpus, but by presenting a large body of samples in an easily navigable and playable way, new sonic combinations emerged. Tools for audio analysis analyse and meter audio differently to the human ear and can, therefore, produce an alternative relational analysis that can act as a rich source of inspiration. This section will give a brief account of the tools used and their potential application. Examples of each tool can be found in *Appendix E*.

XO by XLN Audio[27] is a commercial machine-learning analysis tool designed to analyse large libraries of one-shot drum samples and categorise them. The *Spleeter* corpus [*Appendix B*] was fed into this tool, but most of the sample set content was too long to be compatible with the tool. Despite the incompatibility, it’s clear that *XO’s* focus on playability and flexibility could provide music producers with a new way to navigate large sample libraries. As a commercial product, *XO* limits user customisation to maximise user-friendliness, only presenting sonic analysis in one way. For the purposes of this project, only the time-limited trial version was used to evaluate the software’s capabilities. Videos included in *Appendix E* show a selection of samples from the *Spleeter* corpus and *XO’s* attempts to categorise their timbres. As seen in the videos, *XO* allows users to play preprogrammed drum patterns and swap samples based on AI recommendation. This workflow biases a mainstream ‘digital beat maker’ approach and doesn’t quite support alternative modes of production. It’s probable XO’s nature as a commercial product

26 *AI and Music - Holly Herndon Presents Holly+ Feat. Maria Arnal, Tarta Relena and Matthew Dryhurst*, Live Stream Archive, Sónar 2022 (Barcelona, Spain, 2022), 41:00, https://www.youtube.com/watch?v=Wk6T2WmhuJw.
27 *XO by XLN Audio - Your Universe of Sounds [Announcement Video, April 9th 2019]*, 2019, https://www.youtube.com/watch?v=kYVd01NhAKQ.

has shaped the tool for a mainstream audience, prioritising plug-and-play features and limiting customisability.

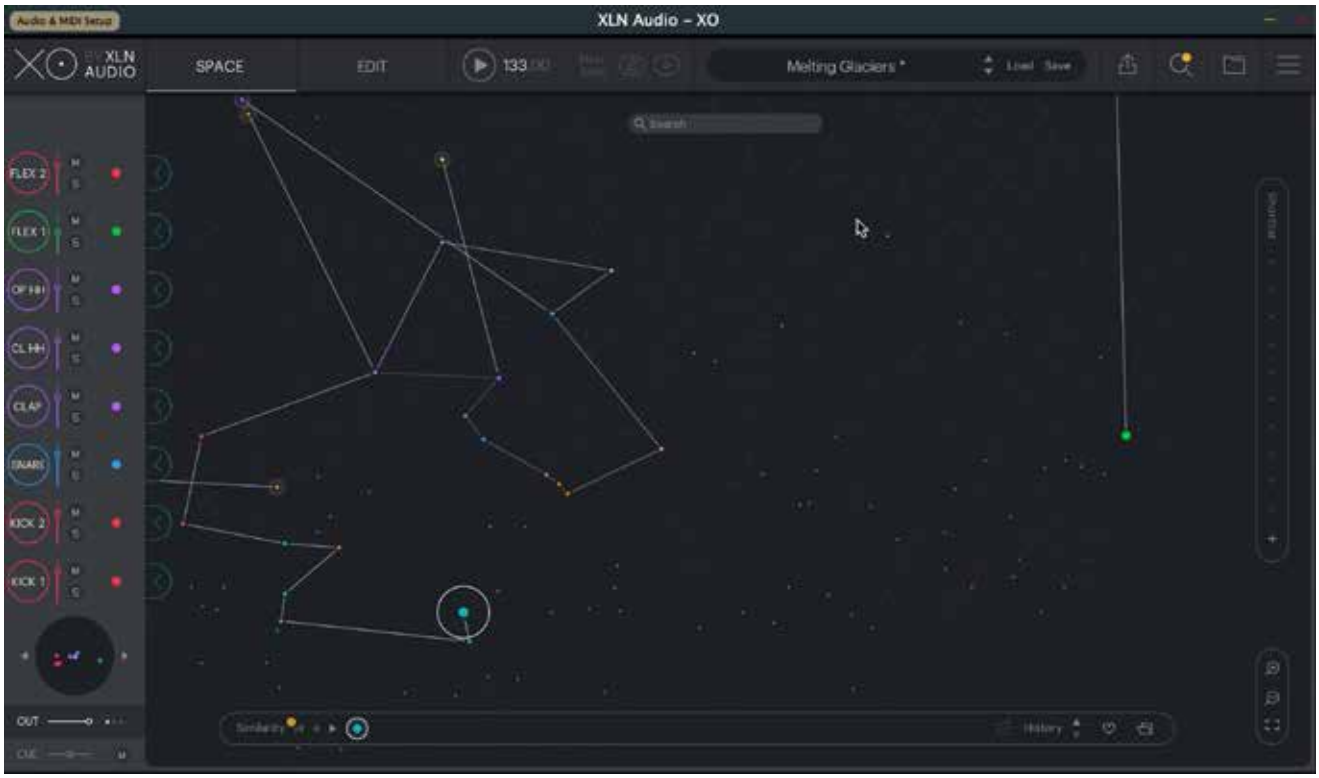

Fig. 1: XO by XLN audio's main screen with Spleeter Corpus samples loaded.

CataRT, developed by IRCAM,[28] is a 'concatenative real-time sound synthesis system' that analyses and plots sound based on several descriptors. The Max/MSP version is freely accessible and highly customisable. Using objects from IRCAM's MuBu and PiPo libraries, the tool is capable of presenting audio datasets in a multitude of contexts with a focus on performability. It presented a flexible mode for visualising a large body of samples. The demonstration videos in *Appendix E* illustrate that this software performs best with short, granular samples with similar timbral or descriptive qualities. When the entire *Spleeter* corpus was fed into CataRT, it was difficult to parse the sonic output, but when the corpus was divided by *Spleeter* output stem-type (Vocal, Drum, Bass, Piano, Other), a clear use case emerged. This tool provided another, customisable, mode for engaging with large sample libraries and corpuses, but it performs best with sounds of similar quality. CataRT rests at an intermediate point between XO and FluCoMa, discussed below. However, as it is built using an extensive library of IRCAM machine-learning audio tools, users can dig into a rich set of features beneath the surface and, with some work, customise the tool to their needs. Regardless, CataRT for Max/MSP is a free, off-the-shelf implementation of these techniques.

28 'CataRT', *Sound Music Movement Interaction - ISMM* (blog), 4 March 2014, https://ismm.ircam.fr/catart/.

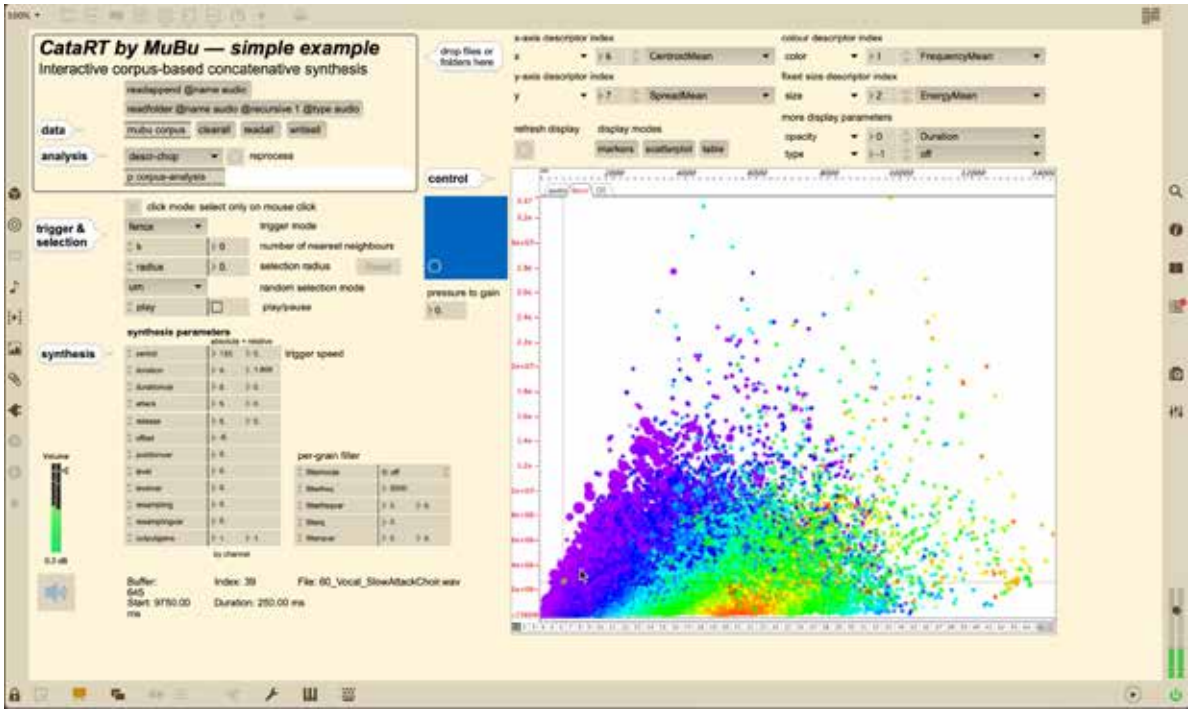

Fig. 2: CataRT by IRCAM for Max/MSP with the Spleeter Corpus

FluCoMa or Fluid Corpus Manipulation is a machine-learning DSP library developed by a team at the University of Huddersfield.[29] It offers flexible, multipurpose tools for employing machine-learning techniques. FluCoMa is a powerful toolset but requires significant time investment into learning in order to achieve usable results. Like CataRT, it's capable of analysing sonic corpuses based on a variety of parameters. While there is no off-the-shelf analysis solution in FluCoMa, the toolkit, particularly the Max/MSP implementations, provides users with the ability to use machine-learning tools in almost any creative context, from more accurate audio analysis and detection to new methods of synthesis and granulation. The videos included in *Appendix E* show an implementation of FluCoMa's sonic corpus exploration capabilities based on a series of tutorials.[30] Engagement with FluCoMa in this project only scratches the surface of the toolkit's potential. Using FluCoMa as a means to process audio with machine learning could constitute an entirely separate research project, and as an in-development tool kit, it will continue to become more flexible and provide technologists with incredibly powerful machine-learning digital signal-processing tools going forward.

29 Pierre Alexandre Tremblay et al., 'Fluid Corpus Manipulation Toolbox', 7 July 2022, https://doi.org/10.5281/zenodo.6834643.

30 James Bradbury, '2D Corpus Exploring', learn.flucoma.org, 10 May 2022, https://learn.flucoma.org/learn/2d-corpus-explorer/.

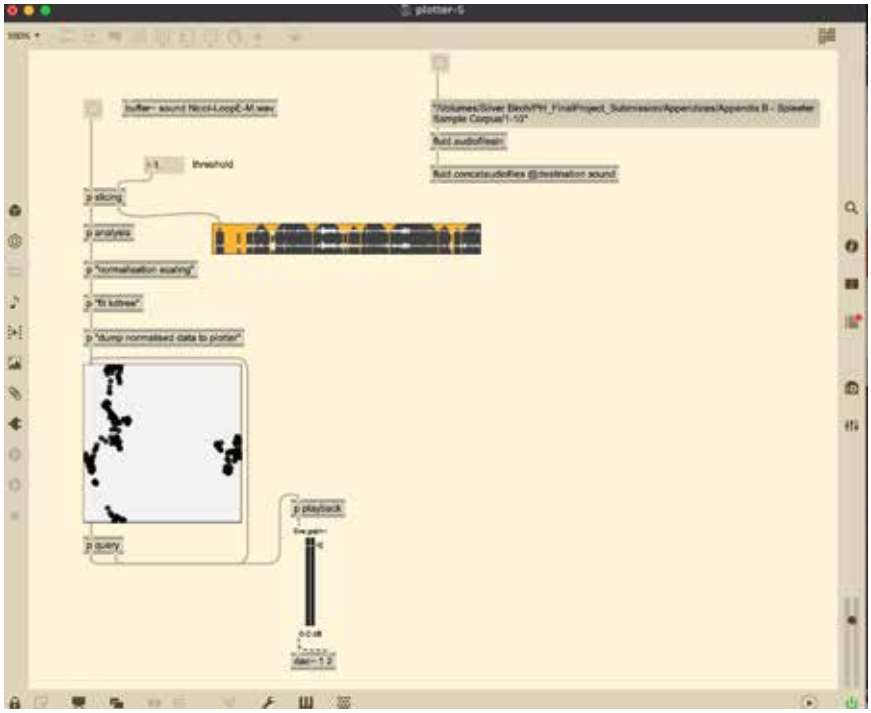

Fig. 3: 2d Corpus exploration patch using FluCoMa in Max/MSP with excerpts of the Spleeter Corpus

To conclude this chapter, it's clear that, while nascent, the field of AI audio processing is varied and compelling. The user-accessible tools discussed in this chapter have the potential to assist in many creative audio workflows and offer a low barrier to entry for anyone interested in the field. Overall, most tools discussed above don't demand deep engagement and could feasibly be used to support existing creative practices without the need for extensive technical knowledge or skill.

Chapter Two: First Voices

Identifying AI audio tools and evaluating their usefulness in a creative process.

In this chapter, the compositions created using AI tools for this project will be discussed. This follows chapter one's survey of currently accessible AI audio processing technology and expands on the possible techniques available to a composer wanting to employ these tools in a creative practice. The pieces that emerged from this project's research will act as the basis for this, presented as both descriptions of process and creative reflection.

First Voices___

First Voices is piece for tape and two performers composed at short notice using AI tools for a PlusMinus Ensemble performance in early July 2022. It was the first finished piece using machine-learning tools for this project.

The piece explores themes of language learning and new life as an AI learns to find its voice for the first time, comparing the process of 'training' an AI with a dataset to learning to speak and articulate thought for the first time.[31] The tape component was created by processing clips of the earliest known recordings from the 1860s[32] with *Holly Plus*. *Holly Plus* produces an uncanny-sounding vocal tone that lends a compelling tonality to the degraded, noisy early recordings, creating musicality out of relatively atonal samples. This tool sounds computerised; it's a warped and uncanny soundalike of Herndon's singing voice. It's both recognisable and extremely alien. It's an incredibly unique sonic quality, and one that demonstrates the compelling quirks and idiosyncrasies of even the most cutting-edge digital tools. Snippets of the processed track were looped, layered and pitched using various digital techniques and arranged into two sections to support the narrative of the vocal component. Finally, the piece was recorded to tape and slowed down slightly to restore an analogue quality to the digital sounds of *Holly Plus*.

The performed vocal component of the piece features two vocalists performing three movements, a recording of which can be found in *Appendix F*. For the first movement, each performer reads lists of randomised phonemes. They recite the list, calling and responding and

31 Sollit, 'Collaborating with Intelligent Machines'.
32 *The Very First Recordings (1859-1879)*, 2019, https://www.youtube.com/watch?v=-0H8Q4QD-cM.

sometimes speaking over each other. The second movement has the performers reciting lists of words in increasing length. These lists were populated from random combinations of common words, but as the words increase in length, meaning emerges from the randomness. This emergent quality evokes the idea of the AI characters learning to speak and form meaning. In the final movement, the AI characters discuss themes of autonomy, identity and sentience, a sudden but evocative contrast from the nascence of the previous movements.

The tape track strives to make a connection between those earliest recordings and the sounds of new audio processing AIs that have a similar low-fidelity quality. The uncanny similarity of sounds produced in these two eras serves as a metaphor for the youth of AI tools, while the final movement draws inspiration loosely from recent headlines discussing an allegedly sentient Google chatbot.[33] The machine learning-generated audio we're hearing today is just as young as those earliest recordings. The tape track can be found in *Appendix F*.

The score is presented as text-only, white text on a black background. The preface discusses digital nativity and the ways in which we perceive information differently depending on the medium: paper or screen. Presenting it as a 'graphic' score with a high degree of improvisation helped to encourage players to perform and embody the voice of the still-growing AIfound in the piece. This was designed to reflect the fact that a machine-learning algorithm is only as good as its dataset – the better the performance, the better the AI. The collected materials ,including recordings, scoreand the tape piece,can be found in *Appendix F*. A full video of the performance by PlusMinus can also be found in *Appendix F*.

This piece was also presented as a work in Unity, taking a performance about the differences between the digital and physical to the digital world to explore the same questions in a different context. See *Appendix G* for a video recording of this piece presented digitally. In this digital presentation, two AI entities are represented by floating, glowing polygons. These shapes are sound-responsive. The piece takes place in a continuous set of identical rooms in a three-by-three grid, allowing users to navigate the repeated, disorienting space.

An Algorithmic Intervention into Field Recordings___

An Algorithmic Intervention into Field Recordings is a set of works that seeks to transform field recordings using a variety of AI tools. The concept emerged while improvising with Google *Magenta DDSP* and trying to push the tool to its limits. A series of field recordings was made,

33 Nitasha Tiku, 'The Google Engineer Who Thinks the Company's AI Has Come to Life', *Washington Post*, 11 June 2022, https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

one in Haymarket, one by the Leith River in Deans Village and one at a rest stop on the French A7 motorway. The field recordings were first cleaned up in Ableton Live 11, deconstructed using *Spleeter* and then morphed using layers of DDSP plug-in processing. Deep listening reveals a call-and-response quality to these pieces, where the DDSP plugins latch onto a particularly tonal element in a recording and transform it into a tonal gesture. It starts to feel like an AI algorithm listening and responding to these recordings, highlighting certain elements and downplaying others based on the capabilities of the technology. The *Spleeter* separations similarly reveal digital artefacts in the recordings and accentuate recording imperfections. Both machine-learning processes unearth abstract sound qualities that manual processing could not. The results are chaotic and imperfect, but the digital transformation of serene soundscapes into digital sonic clusters is compelling. The algorithm is intervening in these recordings of nature and producing something new and previously unheard. The source and processed recordings are included in *Appendix H*.

These field recordings were also presented in Unity and can be found as videos in *Appendix G*. Each processed field recording has been sliced into 30s segments and randomised in a blend container in Wwise. This allows relatively short field recordings to extend over a potentially infinite duration, creating constant variation. Processed single-bird effects have also been placed in the environment for two of the pieces, to introduce more dynamic elements. Presenting these field recordings in a semi-immersive digital setting showcases how pieces like such as this could be developed and presented in the form of an installation or game-piece and demonstrates a path to future work using these techniques.

Miscellaneous Compositional Experiments using the Spleeter Corpus ___

Appendix I contains four compositions created using samples from the *Spleeter* Corpuscorpus. These pieces are experimental and were responsive to the sample library. Every composition in this appendix were was built around pleasing combinations of samples from the library.

‘AI_Explor_1’ was a piece created entirely from samples from the *Appendix A* test set. This sparse percussive piece served as a proof of concept, demonstrating the timbral capabilities of the AI samples while proving that it was possible to construct convincing music from machine-learning tools. This piece demonstrates the various strengths of each *Spleeter* stem. The bass parts of the outputs tend to lack mid- and high-range frequencies, often only outputting the

low and sub-bass frequencies from a track, creating booming, complex bass tones that require layering or processing to reach midrange frequencies.

'AI_Explor_2' combines sounds from *Spleeter*'s vocal outputs with booming taiko percussion. This electronic choir and percussion piece draws on dance music conventions and surfaces the previously described 'underwater', filtered qualities of *Spleeter*'s output. Ticking percussion and other noisy, rhythmic samples are present in this track without heavy processing, creating complex moving parts with a disorienting quality. This piece was the first finished composition using the wider *Spleeter* corpus.

'AI_Explor_3' was an experiment in implementing the machine-learning outputs with non-AI samples. Choir sounds from the *Spleeter* corpus were processed using granulation techniques and combined with club-style drum samples to create an immersive track that draws inspiration from genres such as early-dubstep and 2-step garage and artists such as Burial. Spitfire Labs is used to create the reverberant, euphoric orchestral parts in this piece, which. This piece is arguably the most fully -formed of *Appendix I*, and could viably be heard on a dancefloor. A piece like this proves the capabilities of AI tools to not only be used as the sole source for sonic work but also as a source of inspiration and sampling for a wider practice.

'AI_Explor_4' is unique amongst the tracks in this appendix. This track takes audio from a single, extended *Spleeter* stem and applies processing to create a granular, harsh ambient atmosphere with a climactic, noisy finish. This piece demonstrates *Spleeter*'s capability to recontextualise longer passages of audio and surface unique sonic qualities.

Overall, this *appendix d*emonstrates the timbral flexibility of *Spleeter* as a sound design tool. It's important to note that it is only capable of processing existing audio in a linear fashion and does not necessarily 'respond' to source material. All pieces contained in this *appendix c*ontain parts of the extensive back catalogue of source material used to generate the corpus and, as such, demonstrate an ability to deconstruct and recontextualise existing sounds rather than generating new audio.

'Is Everything Gonna Be Ok?'___

'Is Everything Gonna Be Ok?' was a work-in-progress performance of Nani Porenta's research into game-engine interactivity in live musical performance. A collection of sound from this project was presented as the supporting act for this showcase. This performance took the form of a performance and lecture, with narration discussing the nature of AI creative tools, some

of which was processed via *Holly Plus*. The lecture component explored issues of AI and labour and the ethical use of creative tools, and it briefly described the processes involved in creating some of the work discussed throughout chapter 2. A recording, video and the narration snippets in audio form with *Holly Plus* morphs can be found in *Appendix D*. It's important to note that the narration was casual in tone and makes some relatively sweeping generalisations about the nature of AI creative tools. While not necessarily academically rigorous, the narration takes a broad view on the nature of AI tools and concludes with a call to action for creative practitioners to experiment with machine-learning tools in their creative practice as a new frontier of creative experimentation.

Like any creative audio tool, the machine-learning tools discussed throughout this essay bias a certain sound and workflow. After only a few experiments, it started to become clear what types of sounds the algorithm best responded to in order to produce the most complex, engaging results. With tools such as *Spleeter*, it's easy to conceptualise an emerging reflective creative practice where an artist begins a recursive cycle: creating something, using the creation to generate samples, composing based around the resulting samples and repeating the cycle, potentially creating a lineage of dozens of samples being created in a recursive loop, ultimately resulting in an extensive library of machine-learning samples.

Tools such as *Holly Plus* and *Magenta DDSP* are less recursive, and instead feeling like 'taming a beast'. The unpredictable nature of complex tone-transfer tools required deeper engagement and tweaking of source material to produce more nuanced results. *Magenta DDSP* handles complex sound qualities and timbres relatively poorly, but when pushed to its limits, we see interesting results [See *Appendix H*]. *Holly Plus*'s ability to morph sound snippets into a soundalike of Herndon's own voice is consistently impressive. As a means to imbue complex and pleasing-sounding vocal qualities to nonvocal and even atonal recordings, this tool is unmatched. Tone transfer techniques are, by far, the most abstract and black box-like tools, but the unpredictability of the results remains a constant source of creative inspiration.

With exception of *Spleeter*, it's not likely that these creative tools will supersede or replace any existing creative tools, but they can be considered new additions to the diverse number of experimental sonic-processing techniques available to creative practitioners. *Spleeter*, on the other hand, executes source separation at a fidelity and speed that nonneutral network tools are simply incapable of reproducing. In its current form, the quirks of the tool are exploitable for rich sonic outcomes, but future versions of AI source-separation tools will likely become irreplaceable in postproduction workflows.

Overall, a body of work is presented here that demonstrates an extensive set of applications and use cases for the AI tools discussed. It is hoped that it will encourage readers to engage with these tools on their own terms and explore how nascent machine-learning tools could work in their own creative workflows.

Chapter Three: Algorithm's First Steps

Reflection on machine learning in the compositional process and an ethical AI future.

Despite narrowing the focus of this project to relatively plug-and-play machine-learning interfaces that don't require extensive technical knowledge or large datasets, tools with compelling results and demonstrably broad application still emerged. The continued developments of these still nascent audio tools are likely to provide creators with increasingly powerful toolsets to execute their compositional ideas in the coming decades. As mentioned earlier in this essay, most of the publicly accessible music-generating AI tools that exist today eschew direct creative engagement, instead operating as a proof-of-concept projects based on the results of public and private AI research, but it's easy to imagine this changing fast. It's not difficult to draw a parallel between the machine-learning audio tools emerging today with the image-generation experiments of the mid-2010s. What were once novel experiments and internet oddities are now fully fledged image-generation algorithms capable of re-creating and editing images convincingly.

Fig. 4: A picture of a cat with various machine learning image generation processes applied.

OpenAI's GPT-3 powered *Dall-E 2* image-generation algorithm [See fig. 4] is now starting to be used in major branding projects with graphic designers such as David Rudnick starting to exploringe the power of these technologies as the next frontier of their design practice.[34]

34 Lindsay Howard, 'The Mountains of Idyllwild, but Make It DALL-E', Blog, The Mountains of Idyllwild, but Make It DALL-E, 28 July 2022, https://www.fwb.help/wip/fwb-fest-visuals-dall-e-david-rudnick-laiqa-mohid.

Could we plot the course of AI audio tools on a similar trajectory? We'll likely see a growing ubiquity of AI audio tools, but it's not a solved problem – audio generation required a totally different approach and steeper resource cost to execute; the time domain of audio makes things exponentially more complex. It's difficult to predict just how quickly these tools will develop.

The field of AI and music is growing. Herndon's *Holly Plus*, discussed extensively throughout this essay, is just one of a growing number of practitioner-led AI musical practices. *Holly Plus* itself grew out of Proto, Herndon's 2019 album that used a neural network to respond to snippets of audio sung by her choir. As a work, it centres the chaos and unpredictability of these machine-learning call calls and responses in one of the most interesting albums of that year.[35]

Jennifer Walshe's 'A Late Anthology of Early Music Vol. 1: Ancient to Renaissance' used similar AI training techniques to respond to a corpus of recordings of Walshe's voice and reimagine a body of early western music. The results interrogate the musical canon in a work that is equal parts moving and timbrally disorienting.[36]

Pop band YACHT have deployed AI tools slightly differently. For their 2019 album 'Chain Tripping', they trained bespoke algorithms on their back catalogue to discern patterns in their music and discover whether or not they had a musical 'formula', challenging an AI to make connections in music that the human ear couldn't discern. The resulting album was derived from the band stitching together machine-learning outputs to create a chaotic bricolage of samples and lyrics and forming coherence from machine data. It's a reminder that curatorial intervention is crucial when using these tools – musical meaning-making requires a personal touch.[37]

As a final example, AI band/collective DadaBots is pushing the boundaries of AI. They train neural networks on corpuses of specifically copyrighted metal, jazz and other genres of music and generate continuous livestreams of AI-generated music.[38] These continuous livestream projects are modelled on the 'Youtube Pirate Radio Stations'[39] that continuously livestream music 24/7 and push the limits of AI tools to generate and reinvent familiar sounds in uncanny ways.

Creator-led investigations of AI music tools consistently produce the most engaging results while, as described above, often interrogating the nature of the tools in use. These kinds of

35 McDermott, 'Holly Herndon on Her AI Baby, Reanimating Tupac, and Extracting Voices'.

36 Jennifer Walshe, 'A Late Anthology of Early Music', MILKER CORPORATION, 21 February 2020, http://milker.org/a-late-anthology.

37 Madeleine Brand, 'How YACHT Used A.I. to Make Their Album "Chain Tripping" | Press Play', KCRW, 9 September 2019, https://www.kcrw.com/news/shows/press-play-with-madeleine-brand/using-a-i-to-make-a-music-album/how-yacht-used-a-i-to-make-their-album-chain-tripping.

38 Rich Haridy, 'Meet Dadabots, the AI Death Metal Band Playing Non-Stop on Youtube', New Atlas, 23 April 2019, https://newatlas.com/dadabots-death-metal-neural-network-livestream/59394/.relentless heavy metal music, this latest example of AI-generated creativity could be either glorious ear candy or the aural equivalent of water-boarding. Currently live-streaming on YouTube is a non-stop, algorithmically-generated torrent of technical death…","

39 Jonah E. Bromwich, 'Pirate Radio Stations Explode on YouTube', *The New York Times*, 3 May 2018, sec. Arts, https://www.nytimes.com/2018/05/03/arts/music/youtube-streaming-radio.html.

projects are likely to become more ubiquitous as artists gain access to and the ability to develop and train their own neural networks.

A running theme of this essay has been the high resource cost required to develop AI music tools. The field of AI is dominated by companies with access to unfathomably large datasets – Google, Apple, Facebook, Microsoft, Amazon, IBM, NVIDIA and Tencent – all of which employ data gathered from their other lines of business as fodder for large-scale neural network experiments. With little exception, this research focuses on potential commercial and labour-replacement application – AI for automation. It's crucial that we acknowledge that the AI industry is modelled on what is widely regarded as exploitative labour.[40] Tools such as Amazon's crowdwork platform 'Mechanical Turk' , are relied on by companies of all sizes small and large to crowdsource micro-labour, or the completion of small, repetitive tasks, to train and categorise datasets for neural networks.[41]

It seems quite unlikely that visual artists will be able to develop bespoke image-generation algorithms with small-scale datasets, but the increasing ubiquity of large-scale image tools has come under scrutiny. Artists are quickly realising that the fingerprint of their work is being surfaced by commercial image-generating neural networks, which have scoured image-hosting services for source data.[42] In response to this criticism, these companies rely on the 'black box' nature of the tools for cover, stating that they don't understand the how data for their own commercially available tools is being parsed.[43] Meta (formerly Facebook) has developed an AI chat tool named *BlenderBot* that has faced similar criticism. This chatbot is trained based on open-user input, and the results, darkly, speak for themselves. Facebook has developed a conspiracy-theory laden, Facebook-hating chatbot.[44] How long will it be before legislative intervention is required to step in for reasons of copyright and legality while companies hide behind the 'Black Box' nature of their tools?

40 Moritz Altenried, 'The Platform as Factory: Crowdwork and the Hidden Labour behind Artificial Intelligence', *Capital & Class* 44, no. 2 (2020): 145–58, https://doi.org/10.1177/0309816819899410.

41 Roger Von Laufenberg, 'The Mechanical Turk – or the Invisible, Low-Cost Labour of Automation', VICESSE, 8 February 2022, https://www.vicesse.eu/blog/2022/2/8/the-mechanical-turk-or-the-invisible-low-cost-labour-of-automation.

42 Karla Ortiz [@kortizart], '1/ Hello @midjourney How Goes? Quick Question, Is There a Way One Can Specifically Ask for Their Name to Be in a "Do Not EVER Allow to Be Used in a Prompt" List? Seeing Folks on Your Discord Not Only Use My Name, but Trying to Figure out How to Train Your Ai to Mimic My Work-', Tweet, *Twitter*, 1 August 2022, https://twitter.com/kortizart/status/1553930997590175745.

43 Midjourney [@midjourney], '@samirsns @kortizart @craigmullins3 It Seems to Think Craig Mullins Does "gritty Art" but It Has Extremely Little Knowledge of Anything Else. Often It Learns Those Kinds of Things Just from Word Associations without Any Images. This Is a Pretty New Field and We're Still Learning about What Makes This Stuff Work.', Tweet, *Twitter*, 1 August 2022, https://twitter.com/midjourney/status/1553995668812808192; Midjourney [@midjourney], '@eepoxdraws Current Data Is Broad Scrapes of the Internet. It's Not Known What Is Helpful. The Science Is New, but to Give You Some Idea, It Learns from 250TB but Only Remembers 2GB (125,000x Decimation of Data). Meaning It Mostly Learns High-Level Abstraction and Broad Commonalities.', Tweet, *Twitter*, 1 August 2022, https://twitter.com/midjourney/status/1553998248276152320.

44 Jeff Horwitz [@JeffHorwitz], 'Good Morning to Everyone, Especially the Facebook Http://Blender.Ai Researchers Who Are Going to Have to Rein in Their Facebook-Hating, Election Denying Chatbot Today Https://T.Co/WMRBTkzlyD', Tweet, *Twitter*, 7 August 2022, https://twitter.com/JeffHorwitz/status/1556245316596219904.

In contrast, there is an optimistic side to all of this. The majority of AI audio tools are being trained on significantly smaller scale datasets. *Magenta DDSP* and *Spleeter* both enable users to retrain their algorithms based on very small bodies of work,[45] and many of the examples discussed in this chapter have been developed by artists collaborating with machine-learning experts while drawing from their own back catalogue or using specifically tailored audio corpuses.[46]

It is hoped that the body of work presented in this project demonstrates a multitude of creative and reflective processes for engaging with deep learning AI tools in a compositional context. It is also hoped that the format has inspired other musicians to engage with these easy-to-access tools.

Many industry-standard audio-repair and editing tools are using machine-learning techniques to automate nuanced postproduction processes. The aforementioned *Izotope*, a company that produces audio repair and production tools, has implemented flexible machine-learning tools that analyse and optimise tracks in a variety of ways, including mixing, mastering and balance. These tools have demonstrable potential to save production time and allow practitioners to invest more time in creation, trusting that algorithmic tools will be able to assist them, at least in partpartly, with meeting technical requirements.[47] While, *LANDR*, a service that claims to offer 'AI Music Mastering', is almost 9 years old – these tools have been commercially available for over a decade.[48] It's not hard to imagine that machine-learning algorithms will continue to seep into industry-standard tools as the professional audio industry demands higher and higher workloads.[49]

To briefly talk in the first person, I have found this project highly rewarding. By turning to this new frontier of digital musical tools, I have found my own back catalogue recontextualised and revitalised. These tools will certainly form part of my ongoing creative practice far into the future, and it has been consistently exciting to make work with these unpredictable and often undecipherable tools during the course of this project. Almost every part of my creative practice has found itself pushed and challenged, forced to reinvent and respond to the use of these tools. It has been both academically and creatively satisfying to develop work with and about AI and machine-learning tools, and I hope that my enthusiasm and optimism for these nascent toolkits is

45 Engel et al., 'DDSP'.
46 Walshe, 'A Late Anthology of Early Music'; McDermott, 'Holly Herndon on Her AI Baby, Reanimating Tupac, and Extracting Voices'.
47 Spratley, Beck, and Cohn, 'A Unified Neural Architecture for Instrumental Audio Tasks'.
48 Tyler Hayes, 'Why The Music Industry's Next Big Disruption Is In The Recording Studio', Fast Company, 2 July 2014, https://www.fastcompany.com/3032642/why-the-music-industrys-next-big-disruption-is-in-the-recording-studio.
49 Spratley, Beck, and Cohn, 'A Unified Neural Architecture for Instrumental Audio Tasks'.

clear from the work produced. The work presented here only scratches the surface of the impact of these tools on my workflow – I have improved equally both as an academic, able to engage critically with a field and set of tools and, as a composer and practitioner, making work that challenges and excites. On those grounds alone, I deem this investigation a personal success.

In conclusion, it's clear that while the still nascent AI industry is dominated by those with the computational resources to invest in research, diverse and compelling creative tools, use cases and creative practices are emerging. Increasing numbers of tools misguidedly seek to replace the artist, or at least act as a surrogate for the creative process, but it is unlikely that these will dominate the field.[50] As the technical investment required to solo-develop machine-learning tools reduces, there's incredible room for creatives to respond and train their own algorithms with data based on their own practice. Many artists are already at the forefront of applying these technologies in existing creative workflows, using them to prompt and derive inspiration for large, compelling bodies of work, and this trend will only increase in the future.

50 Altenried, 'The Platform as Factory'.

This Page Left Blank Intentionally

Bibliography

Abadi, Martin, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. ‘TensorFlow: A System for Large-Scale Machine Learning’. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–83, 2016. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.

AI and Music - *Holly Herndon Presents Holly+ Feat. Maria Arnal, Tarta Relena and Matthew Dryhurst.* Live Stream Archive. Sónar 2022. Barcelona, Spain, 2022. https://www.youtube.com/watch?v=Wk6T2WmhuJw.

Altenried, Moritz. ‘The Platform as Factory: Crowdwork and the Hidden Labour behind Artificial Intelligence’. *Capital & Class 44*, no. 2 (2020): 145–58. https://doi.org/10.1177/0309816819899410.

Bawiec, David. ‘IZotope and Machine Learning: Speeding Up Your Workflow with Assistive Audio Technology’. iZotope, 9 January 2020. https://www.izotope.com/en/learn/speed-up-your-workflow-with-assistive-audio-technology.html.

Bradbury, James. ‘2D Corpus Exploring’. learn.flucoma.org, 10 May 2022. https://learn.flucoma.org/learn/2d-corpus-explorer/.

Braguinski, Nikita. *Mathematical Music: From Antiquity to Music AI*. Oxford, England ; Routledge, 2022.

Brand, Madeleine. ‘How YACHT Used A.I. to Make Their Album “Chain Tripping” | Press Play’. KCRW, 9 September 2019. https://www.kcrw.com/news/shows/press-play-with-madeleine-brand/using-a-i-to-make-a-music-album/how-yacht-used-a-i-to-make-their-album-chain-tripping.

Bromwich, Jonah E. ‘Pirate Radio Stations Explode on YouTube’. The New York Times, 3 May 2018, sec. Arts. https://www.nytimes.com/2018/05/03/arts/music/youtube-streaming-radio.html.

Brown, Sara. ‘The Hidden Work Created by Artificial Intelligence Programs’. MIT Sloan, 4 May 2022. https://mitsloan.mit.edu/ideas-made-to-matter/hidden-work-created-artificial-intelligence-programs.

Sound Music Movement Interaction - ISMM. ‘CataRT’, 4 March 2014. https://ismm.ircam.fr/catart/.

Dobrev, Lyubomir. ‘AudiaLab Synapse Drums Creates Drum Samples with Artificial Intelligence’. gearnews.com, 10 February 2022. https://www.gearnews.com/audialab-synapse-drums-creates-drum-samples-with-artificial-intelligence/.

Engel, Jesse, Hanoi Hantrakul, Chenjie Gu, and Adam Roberts. ‘DDSP: Differentiable Digital Signal Processing’. Magenta, 15 January 2020. https://magenta.tensorflow.org/ddsp.

Haridy, Rich. ‘Meet Dadabots, the AI Death Metal Band Playing Non-Stop on Youtube’. New

Bibliography Cont.

Atlas, 23 April 2019. https://newatlas.com/dadabots-death-metal-neural-network-lives-tream/59394/.

Hayes, Tyler. 'Why The Music Industry's Next Big Disruption Is In The Recording Studio'. Fast Company, 2 July 2014. https://www.fastcompany.com/3032642/why-the-music-industrys-next-big-disruption-is-in-the-recording-studio.

Herndon, Holly. 'Holly+ 💪 👭🗣️💪', 13 July 2021. https://holly.mirror.xyz/54ds2IiOnvthjGFkokF-CoaI4EabytH9xjAYy1irHy94.

Howard, Lindsay. 'The Mountains of Idyllwild, but Make It DALL-E'. Blog. Friends With Benefits/FWB, 28 July 2022. https://www.fwb.help/wip/fwb-fest-visuals-dall-e-david-rudnick-laiqa-mohid.

Huang, Cheng-Zhi Anna, Tim Cooijmans, Monica Dinculescu, Adam Roberts, and Curtis Hawthorne. 'Coconet: The ML Model behind Today's Bach Doodle'. Blog. Magenta, 20 March 2019. https://magenta.tensorflow.org/coconet.

Jeff Horwitz [@JeffHorwitz]. 'Good Morning to Everyone, Especially the Facebook Http://Blender.Ai Researchers Who Are Going to Have to Rein in Their Facebook-Hating, Election Denying Chatbot Today Https://T.Co/WMRBTkzlyD'. Tweet. Twitter, 7 August 2022. https://twitter.com/JeffHorwitz/status/1556245316596219904.

Karla Ortiz [@kortizart]. '1/ Hello @midjourney How Goes? Quick Question, Is There a Way One Can Specifically Ask for Their Name to Be in a "Do Not EVER Allow to Be Used in a Prompt" List? Seeing Folks on Your Discord Not Only Use My Name, but Trying to Figure out How to Train Your Ai to Mimic My Work-'. Tweet. Twitter, 1 August 2022. https://twitter.com/kortizart/status/1553930997590175745.

Knotts, Shelly, and Nick Collins. *'A Survey on the Uptake of Music AI Software'.* Proceedings of the International Conference on New Interfaces for Musical Expression. Presented at the International Conference on New Interfaces for Musical Expression, Birmingham, UK, 1 June 2020. https://doi.org/10.5281/zenodo.4813499.

McDermott, Emily. 'Holly Herndon on Her AI Baby, Reanimating Tupac, and Extracting Voices'. ARTnews.Com (blog), 7 January 2020. https://www.artnews.com/art-in-america/interviews/holly-herndon-emily-mcdermott-spawn-ai-1202674301/.

McLeavey Payne, Christine, Justin Jay Wang, Nicholas Benson, and Eric Sigler. 'MuseNet'. OpenAI, 25 April 2019. https://openai.com/blog/musenet/.

Midjourney [@midjourney]. '@eepoxdraws Current Data Is Broad Scrapes of the Internet. It's Not Known What Is Helpful. The Science Is New, but to Give You Some Idea, It Learns from 250TB but Only Remembers 2GB (125,000x Decimation of Data). Meaning It Mostly Learns High-Level Abstraction and Broad Commonalities.' Tweet. Twitter, 1 August 2022. https://twitter.com/midjourney/status/1553998248276152320.

———. '@samirsns @kortizart @craigmullins3 It Seems to Think Craig Mullins Does "gritty Art"

Bibliography Cont.

but It Has Extremely Little Knowledge of Anything Else. Often It Learns Those Kinds of Things Just from Word Associations without Any Images. This Is a Pretty New Field and We're Still Learning about What Makes This Stuff Work.' Tweet. Twitter, 1 August 2022. https://twitter.com/midjourney/status/1553995668812808192.

Mordvintsev, Alexander, Christopher Olah, and Mike Tyka. 'DeepDream - a Code Example for Visualizing Neural Networks'. Google AI Blog (blog), 1 July 2015. http://ai.googleblog.com/2015/07/deepdream-code-example-for-visualizing.html.

Moussallam, Manuel. 'Releasing Spleeter: Deezer R&D Source Separation Engine'. Medium, 3 February 2020. https://deezer.io/releasing-spleeter-deezer-r-d-source-separation-engine-2b88985e797e.

Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 'Hierarchical Text-Conditional Image Generation with CLIP Latents'. arXiv, 12 April 2022. https://doi.org/10.48550/arXiv.2204.06125.

Sollit, Lucy. 'Collaborating with Intelligent Machines'. Intersections: Art and Digital Creativity in the UK (blog), 27 April 2017. https://medium.com/intersections-arts-and-digital-culture-in-the-uk/collaborating-with-intelligent-machines-cb5ecf32c98d.

Spratley, Steven, Daniel Beck, and Trevor Cohn. 'A Unified Neural Architecture for Instrumental Audio Tasks'. arXiv, 28 February 2019. https://doi.org/10.48550/arXiv.1903.00142.

The Very First Recordings (1859-1879), 2019. https://www.youtube.com/watch?v=-0H8Q4QD-cM.

Tiku, Nitasha. 'The Google Engineer Who Thinks the Company's AI Has Come to Life'. Washington Post, 11 June 2022. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

Tremblay, Pierre Alexandre, Owen Green, Gerard Roma, James Bradbury, Ted Moore, Jacob Hart, and Alex Harker. 'Fluid Corpus Manipulation Toolbox', 7 July 2022. https://doi.org/10.5281/zenodo.6834643.

Audacity Team Wikipedia. 'Vocal Removal Plug-Ins - Audacity Wiki'. Wikipedia, 13 October 2020. https://wiki.audacityteam.org/wiki/Vocal_Removal_Plug-ins.

Von Laufenberg, Roger. 'The Mechanical Turk – or the Invisible, Low-Cost Labour of Automation'. VICESSE, 8 February 2022. https://www.vicesse.eu/blog/2022/2/8/the-mechanical-turk-or-the-invisible-low-cost-labour-of-automation.

Walshe, Jennifer. 'A Late Anthology of Early Music'. MILKER CORPORATION, 21 February 2020. http://milker.org/a-late-anthology.

Bibliography Cont.

XO by XLN Audio - Your Universe of Sounds [Announcement Video, April 9th 2019], 2019. https://www.youtube.com/watch?v=kYVd01NhAKQ.

Appendices

Appendix A

The folder for *Appendix A* contains a set of musical sketches and the resulting stems and samples from processing using *Spleeter*. The *Spleeter* implementation for this portion of the project used mvsep.com, a free, online, source separation algorithm website that allows users to upload tracks and produces stems using a variety of source separation processes.

Appendix B

The folder for *Appendix B* contains the resulting sample corpus created by batch processing 150 demos from 2018 to 2022 via the desktop implementation of *Spleeter*. There are over 970 samples, each loosely labelled. The result is a body of samples that are incredibly varied, used making a large library of old music.

Appendix C

The folder for *Appendix C* contains a diverse variety of audio experiments using DDSP tools such as *Magenta DDSP* and Holly Herndon's *Holly Plus*. Chapter 1 contains explanations and comparisons of these results.

Source files labelled 'Source A', 'Source B', 'Source C', and 'Source D', were processed using a variety of tonal transfer parameters to morph simple and relatively complex audio. *Magenta DDSP* does this process in real time, while *Holly Plus* is a non-real-time tool in its current form.

Appendices Cont.

Appendix D

The folder for *Appendix D* contains material from a live performance of project material titled ‘Is Everything Gonna Be Ok?’ on Tuesday 2nd August 2022. This performance was the supporting act for Nani Porenta’s interactive music project of the same name. Included in the *appendix i*s a full audio recording of the performance, narration segments and snippets processed via *Holly Plus*, a collection of photos from the event and a video recording.

It’s important to note that the narration contained within this piece makes generalisations about the nature of creative AI and is not academically rigourous. The piece was put together last minute and was designed to embody a slightly provocative perspective on AI tools for a general audience. It references many of the themes touched on in this essay.

Appendix E

The folder for *Appendix E* contains screengrabs and snapshots of various analytical tools used in the course of this project to categorise the *Spleeter* corpus found in *Appendix B*. The engagement with these tools is relatively surface-level; deeper exploration was intended but, due to time constraints, never completed.

XO by XLN audio, IRCAM’s CataRT and FluCoMa, all present flexible methodologies to engage with a body of samples and find ways to surface unique combinations of sounds with the intention to spark interest and provide inspiration to artists creating using these tools. It’s important to note that the full-featured but limited time trial of XO by XLN audio was used, while CataRT’s and FluCoMa were explored via implementations for Max/MSP.

Appendix F

The folder for *Appendix F* contains collected material from First Voices, a composition for 2 voices and tape. The folder contains tape piece, score drafts and audiovisual recordings of a live performance of the piece.

Appendices Cont.

Appendix G

The folder for *Appendix G* contains screen recordings of various elements from this project presented in Unity. Due to time constraints, these pieces have not been built for runtime and included alongside this submission. Despite this, these presentations demonstrate an alternative way to engage with works using and centring AI tools.

Appendix H

The folder for *Appendix H* contains collected field recordings and the AI processed versions titled ‘An Algorithmic intervention into Field Recordings’.

Appendix I

The folder for *Appendix I* contains a selection of musical examples created using a variety of sonic AI sources.

Philly Holmes - 2022

to the moon & stars & back, forever.

THE UNIVERSITY *of* EDINBURGH
Edinburgh College of Art

Image courtesy of OpenAI Dall-E 2 (August 2022)