Artificial Intelligence Within the Creative Process of Contemporary Classical Music

ROBERT LAIDLOW

PhD 2022

Artificial Intelligence Within the Creative Process of Contemporary Classical Music

Robert Laidlow

A thesis submitted in partial fulfilment of the requirements of the Royal Northern College of Music and Manchester Metropolitan University for the degree of Doctor of Philosophy

Royal Northern College of Music and the Faculty of Arts and Humanities, Manchester Metropolitan University

In partnership with the North-West Consortium Doctoral Training Partnership and the BBC Philharmonic Orchestra

July 2022

Table of Contents

Contents

Portfolio Contents

Original Compositions

Turing Test // Prelude

4 minutes

2019

Solo keyboard instrument

Premiered in March 2019 by Mahan Esfahani at the Barbican Centre, as part of the Barbican Centre event ‘Goedel Escher Bach: The Eternal Golden Braid’.

Provided recording made by Joseph Havlat.

Three Entistatios

12 minutes

2019

12-part chamber ensemble

Premiered in June 2019 by the RNCM New Ensemble at the Royal Northern College of Music, as part of the RNCM’s Future Music Festival. Conducted by Jack Sheen.

Provided recording made by the BBC Philharmonic Orchestra.

Alter

12 minutes

2019

Mezzo-soprano and chamber ensemble

Premiered in November 2019 by Marta Fontanals-Simmons and the Britten Sinfonia. Conducted by William Cole. Commissioned by the Barbican Centre for the PRiSM event ‘Ada Lovelace: Imagining The Analytical Engine’

Provided recording from premiere.

Rose Green

8 minutes

2021

Fixed media

Premiered online in June 2021 as part of the Unsupervised concert series.

No score provided – audio-visual recording only.

Disc Fragments

13 minutes

2021

Tenor and synthesizer

Premiered in Oxford in November 2021 by Louis Watkins and Joseph Havlat. Written as part of the ‘Sound of Contagion’ research project, supported by The Oxford Research Centre for Humanities and the University of the Arts, Berlin

Provided recording from premiere.

Gravity

21 minutes

2021

String quartet

Premiered in June 2021 at the 1901 Arts Club, London.

Commissioned by the Echea string quartet with support from the RVW Trust and the Nicholas Boas Charitable Trust

Provided studio recording made by the Echea Quartet, supported by RNCM PRiSM.

Chromodynamics

9 minutes

2020-21

10-part chamber ensemble

Premiered in October 2021 by Ensemble 10/10, conducted by Robert Ames. Commissioned by the Royal Philharmonic Society and the Liverpool Philharmonic Orchestra
Provided recording from premiere.

Warp
12 minutes
2021
Piano solo and orchestra
Premiered in November 2021 by the BBC Philharmonic Orchestra and Joseph Havlat. Conducted by Vimbayi Kaziboni. Commissioned by the BBC Philharmonic Orchestra
Provided recording from premiere.

Silicon
38 minutes
2020-2022
Orchestra and electronics
Premiered in October 2022 by the BBC Philharmonic Orchestra, conducted by Vimbayi Kaziboni.

Appendix Scores

Three AI Folk Songs
4 minutes
2021
Violin solo
Premiered in August 2021 at the British Embassy, Beijing.
Commissioned by the Alan Turing Institute for AI.
Score only – no recording provided.

Silicon Body PART FOR DDSP SYNTHESIZER

Appendix Audio Tracks

Alter electronics tracks (collated into one file)

Imitation from *Disc Fragments* electronics track

Silicon electronics tracks (collated into one file)

Acknowledgements

I am hugely grateful to my supervisory team, who have supported me in every way throughout the process of this PhD. This was especially true over the period of lockdowns during COVID-19. They are: Emily Howard, David de Roure, Simon Webb, and Marcus du Sautoy. Thank you for believing in me, challenging me, and steering me the right directions.

I am equally grateful to my extremely generous collaborators, who lent me their time, expertise, and AI algorithms for me to experiment with throughout these four years. They include Christine Payne, Jesse Engel, Christopher Melen, Parag Mital, Omar Percha, Hongshuo Fan, and many others.

I am also very grateful to the Royal Northern College of Music, the North West Consortium Doctoral Training Partnership, the RNCM Research Department, RNCM PRiSM, and to the University of Oxford for providing financial support to so many of the projects discussed in this commentary.

In addition, my sincere thanks go to:

Those who have performed my music, including my industry partner the BBC Philharmonic Orchestra, the Britten Sinfonia, the RNCM New Ensemble, the Echea Quartet, the Elias Quartet, the Royal Liverpool Philharmonic Orchestra, Bandwidth Ensemble, Emily Earl, Mahan Esfahani, Joseph Havlat, Marta Fontanals-Simmons, Jack Sheen, William Cole, and Vimbayi Kaziboni.

My collaborators on the 'Sound of Contagion' project Chelsea Haith, Wenzel Mehnert, Sara Laubscher, and the Oxford-UdK Partnership.

Those who look at and listened to my music and gave me comments, including Isabella Gellis, Robin Haigh, Geoffrey King, Athanasia Kontou, Rafael Marino Arcaro, and Andrew Norman.

And, of course, to my incredible wife Natalie who has not only supported me in every way over the last few years but has also shared on so many occasions her vast expertise in machine learning with me.

Abstract

This submission consists of nine pieces of original music in addition to a reflective and critical commentary. With one exception, these pieces are each for live performance, written for ensembles and soloists of various descriptions. The exception is an audio-visual work for fixed media.

These pieces were written as part of my practice-based research PhD and concern the relationship between artificial intelligence and my compositional process. They outline the development of my compositional practice, resulting in the piece *Silicon* for orchestra and electronics which forms a major part of this submission.

The commentary details the algorithms used in the creation of this music, and the aesthetic concerns I developed through working with artificial intelligence. These include the relationship between future and past, authorship, authenticity, musical structuralism, and agency, amongst others. It also describes methods and techniques relating to specific musical elements I developed through working with AI which have had a significant impact on my work.

This research builds upon the areas of research related to my own, especially contemporary classical music, creativity and its relationship to artificial intelligence, machine learning, and algorithmic music practice. It is intended to contribute to the growing field of artistic research that exists within and between these areas.

List of Figures

Commentary Overview

This commentary consists of 5 chapters examining 9 pieces of music. Together they form a technical and musical study of my work for this PhD utilising artificial intelligence (AI) as a creative tool.

The commentary is not intended to be aimed at machine learning specialists, since it exists to support and complement a portfolio of original musical compositions. As such, I have limited my explanation of AI terms to an absolute minimum, only including technical information where that technical information has directly influenced or inspired my compositional process. The commentary includes a glossary at the end which defines more fully terms relating to AI. **Words marked in bold are defined and discussed in the glossary.**

In the context of *this commentary*, I have generally treated the terms AI and machine learning as interchangeable, because the distinction is not relevant to my artistic work. Similarly, network and algorithm are broadly synonymous, and the word model is used to describe the result of an AI algorithm from which I usually generate music, text, or another type of data.

In Chapter 1, the introduction, I will give an outline of the relevant research areas pertinent to my project. There are four main areas: contemporary classical music, creativity and AI, algorithmic music, and musical machine learning. In this chapter, I set out the boundaries of my project and what I am specifically interested in exploring within the vast field of AI and music.

Chapter 2 is a technical study of pieces of music composed using AI in the compositional process. This chapter is not intended to be a complete musical analysis of these works, but rather a specific illumination of the role of AI relating to their composition. The pieces in these chapters can be imagined as prototypes where I am probing the utility and function of different AI algorithms for my own compositional voice. Chapter 2 concludes with my description of the wider aesthetic questions raised by these algorithms.

Chapter 2 discusses five pieces: *Turing Test // Prelude, Three Entistatios, Alter, Rose Green,* and *Disc Fragments.* It describes the use of the AI algorithms 'MuseNet', 'Clara', 'WaveNet', 'SampleRNN', 'LakhNES', 'Synth1GAN', 'FakeYou', 'MarI/O', 'Gym', 'GPT-2', and two custom Text-RNNs created for *Alter*.

Chapter 3 shows the development of my compositional practice following the experimentation of Chapter 2. It examines the pieces *Chromodynamics, Gravity,* and *Warp.*

Chapter 4 is a musical and technical analysis of the piece *Silicon* for orchestra and artificial intelligence. The most intuitive way to read this commentary might be to imagine it as a funnel towards Chapter 4, because this piece develops both aesthetic and technical considerations raised throughout the rest of the commentary. Since *Silicon* is the longest part of my portfolio, at around 35 minutes, this chapter consists of a substantial analysis and reflection. It also describes the use of 'MuseNet', 'DDSP', 'FolkRNN', 'SampleRNN', and 'RAVE' within the compositional process of *Silicon*.

Chapter 5 is the conclusion to the commentary. I reflect on what AI means to me as an artist, how I might generalise what I have learned from using it, and what effect working with AI and computer scientists throughout this PhD has had on my music.

This commentary also includes four appendices. Appendices 1 and 2 cover further algorithms and pieces I created during my experimentation period and could be read either as an addition to Chapter 2 or to Chapter 4. Appendix 3 is a transcript of text recited by AI during the performance of *Alter*. Appendix 4 is further information on *Gravity*.

1: Introduction

1.1 Aims of this Commentary

This commentary accompanies a portfolio of nine compositions written between 2018 and 2022. These compositions range in length, style, and forces, but are all aligned in their contribution to my understanding of how artificial intelligence (AI) relates to musical composition, and ultimately what AI is - or can be - in the context of my own artistic and research development.

My work has focussed on the integration of AI into a compositional process for live musicians, often instrumental. This is partly due to my background as a composer who has always been interested in writing for live performers, and due to my relationship with the BBC Philharmonic Orchestra, the industry partner for this PhD.

AI is a vast area of research, and any project can only engage with a very small part of this field, especially given the rate of technological advancement. Broadly, my aims were to explore AI as a compositional tool, both to generate musical material and to challenge my understanding of composition more generally, and to use music as a means of exploring some elements of the politics of AI.

The commentary exists to provide context and explanation for aspects of the portfolio that are not immediately apparent from the provided scores and/or recordings. Especially, it exists to lift the curtain on where and how AI was used in the compositional process. This is not always obvious from the work alone, particularly when working with AI that is integrated into, and therefore hidden within, the fabric of the music itself.

This introduction will provide a background to my practice as a composer, before situating my work in relation to four fields which closely relate to the work in this portfolio: contemporary music, AI and creativity, musical machine learning (the computational exercise of creating ‘realistic’ music using machine learning), and algorithmic music (the musical exercise of composing with algorithmic and computational processes).

1.2 Background

My compositional work prior to this project focussed on two main areas, scientific collaboration and jazz-infused composition, both of which influence the work undertaken here. I have explored the elegance and beauty of mathematics and physics through collaborations with scientists on several projects, including *Half-life* (2016) for saxophone and nuclear radiation detector, and *Invisible Horizon* (2017) for horn. Similarly I have approached AI in this project, not as an expert in programming or statistics, but as someone who wanted to discover what this field could offer my compositional work. Collaboration with computer scientists has been central to many of the pieces in this portfolio.

My musical background as a jazz performer has likewise influenced my work. Pieces written before this PhD, such as *Lines between* (2018) and *Unsteady Ground* (2018), directly and audibly utilise bebop melodic writing, free improvisation, and other elements I associated with jazz as part of their compositional language. More subtly, jazz harmony has consistently influenced the way I plan harmonic and tonal structures. This can still be heard in many of the works in this portfolio, amongst other interests and passions I have discovered along the way. My experimentation with AI was certainly encouraged by my love of jazz; particularly, I was drawn to the lack of fine control inherent in using an AI to generate musical material for a piece. I see this as similar to incorporating types of improvisation, which also comes with an element uncertainty for the composer.

1.3 Contemporary Music

Over the course of this PhD, I have come to believe that the relationship between people and technology is one of the most important social issues of today. I am indebted to the following composers as examples of how to practically mobilise concert music as a staging ground for discussing, disrupting, and highlighting this relationship. While not all of them are interested in extra-musical arguments in their work, they have all contributed in some way to my understanding of how to communicate my research through music.

My interest in form as primary musical material began with a fascination with teleological music, especially the abstract symphonic music of Beethoven's middle and late periods. My concern with form led me to study modern orchestral music focussing on this area. This includes the music of Andrew Norman, which treats musical form as a distinct phenomenon to be manipulated and developed (*Play* 2013 & *Sustain* 2018). Composers specifically interested in recontextualising musical time as a spatial phenomenon have also been a strong influence on my composition. To give three examples, in *Moult* (2019), Clara Iannotta presents previous iterations of music simultaneously to recontextualise and spatialise time, and both the *Become* trilogy by John Luther Adams (*Become River* 2010; *Become Ocean* 2014; *Become Desert* 2018) and Harrison Birtwistle's *Deep Time* (2016) utilise spatialisation within the orchestra to translate geological time into sound.

Both Norman and Luther Adams additionally use the orchestra as a reflective tool to model extra-musical relationships, with Norman exploring control and power dynamics (*Switch* 2015) and Luther Adams aiming to create a contemplative area for an audience to consider their personal relationship with the natural world. This has deeply influenced the development of my understanding of the both the orchestra itself and the act of orchestral performance as a reflective space and 'rehearsal' for wider social questions posed by AI. Work by composers such as George Lewis (*Minds in Flux* 2021) and Simon Steen-Andersen (e.g., *TRIO* 2019), have contributed directly to my understanding of technology can be used within ensemble music. More broadly, my orchestral writing has also been influenced by composers such as Anna Thovaldsdottir (Aeriality 2011), Edmund Finnis (*Shades Lengthen* 2015), David Sawer (*The Greatest Happiness Principle* 1997), and John Adams (particularly earlier works such as *Grand Pianolo Music* 1981 and *Harmonium* 1981), all of whom have found methods to retain their strong compositional voice while writing for the ensemble.

I am deeply interested in the use of patterns and systems as compositional devices and have been profoundly influenced by music that actively exploits the relationship between system and intuition. The music of Emily Howard (*Torus* 2016 and *Afference* 2014), Thomas Adès (*Concentric Paths* 2005 and *Piano Quintet* 2001), Oliver Knussen (*O Hototogisu!* 2017)and Alban Berg (*Violin concerto* 1935) have served as inspiration on balancing melodic, harmonic, and structural patterns with music intuitively composed; or, on the relationship between material that exists *inside* the grammar of a musical work and material that comes from *outside* that grammar. Similarly, I have found myself influenced by creative orchestrations of abstract systematic music, such as Ensemble Modern's orchestrations of Nancarrow's studies for player piano on the album *As Fast As Possible*.

Adès has also contributed to my understanding of reference as musical material (*Asyla* 1997 and *Darkness visible* 1992), alongside other contemporary composers such as Caroline Shaw (*Plan & Elevation* 2019), Michael Gordon (*Rewriting Beethoven's Seventh Symphony* 2006), Sky Macklay (*Many Many Cadences* 2016), and Oliver Leith (*Honey Siren* 2020). These works deal with reference both ironically and sincerely, informing my approach to the possibilities of self-aware reference and encouraging me to broaden my natural compositional style through examination of other musical genres.

My understanding of intuition in the compositional process is influenced by composers engaged with the possibilities of a given instrument or performer, which I find a productive counterbalance to notions of form and time which can be considered more abstract. Rebecca Saunders and Salvatore Sciarrino are both major influences, especially Saunders' *Skin* (2016) and *to an utterance* (2020) and Sciarrino's *la bocca, i piedo, il suono* (2001), which each involve relatively large ensembles. So too is Ligeti (e.g., *Cello Concerto* 1966 and the string quartets), whose music would also fit in any of the other areas of influence I have described above. In all of these works, ensemble writing is derived from specific performative and embodied possibilities relating directly to a solo instrument or voice and the music would fundamentally not make sense if it were transcribed for a different ensemble; it is wholly idiomatic in every sense. I have found the idea of balancing music that is wholly idiomatic in this way with music driven by form and time to be very rewarding.

1.4 Creativity and AI

Throughout my PhD, the most common questions asked to me by audiences at concerts and peers at conferences are: 'does the AI count as a composer' and 'which bits did the AI write?'. Working with machine learning has caused me to reconsider my notions of what constitutes creativity, a creative act, and the creative process behind making music. Although I came to these issues through machine learning, they are not specific to machine learning and here I show briefly how my understanding of this area is informed by work investigating distributed or alternative creative frameworks.

The issue of creative frameworks, particularly those that try to contextualise the relationship between performer and composer, has been passionately debated in recent years alongside the deconstruction of the mythological composer-as-sole-genius (Whittall 2017). The relationship between a composer and a data scientist or algorithm is very different to that with a performer, but some general principles can be observed. Jennifer Walshe (2016) notes in her description of 'The New Discipline' that composers wishing to make use of drama do not 'have aspirations to start a theatre group – they simply need to bring the tools of the director or choreographer to bear on compositional problems'. I argue the same should be true when working with machine learning – a composer should not need to become a programmer.

Cassandra Miller (2018) and Zubin Kanga (2014) both describe collaborative processes that involve the sending of material from one collaborator to another, and the subsequent analysis or editing of that material according to each person's expertise or subjective sensibilities. This is an approach that I have also taken, replacing the role of one collaborator with a machine learning algorithm.

Jennifer Torrence (2018) has proposed a model that treats performers as a 'deviser' where 'both parties contribute to creative and practical decision making', an aim shared with my methodology. However, Torrence suggests that every individual step includes 'all participating artists regardless of the artist's expertise', which cannot be true in a process that involves working with an algorithm, because one algorithm cannot (yet) contribute to every step in a process. Rather than describing a work as created by one or more 'artists', I prefer to consider both human creators and algorithms to be links in a chain of creative acts, in a similar fashion to Juliet Fraser's (2019) recasting of composer and performer to 'agent[s] in the process of creating the work'.

1.5 Musical Machine Learning

Machine learning is a sub-field of AI that is widely used across all disciplines, including applications of AI in music. My research has been very closely intertwined, at least on a technological level, with what I have termed the 'musical machine learning' field: the field of research advanced by computer scientists where the goal is to create computational methods that analyse, categorise or generate music.

While not all artificial intelligence research concerns machine learning (see Fernández & Vico 2013 for a detailed overview), in recent years machine learning has become the de-facto approach for AI research in the field of music. *Machine learning* specialises in algorithms that learn through experience: a **dataset** is provided (the 'training data') and the algorithm learns by **training** on this data. It will then complete a task, and if it does not complete the task according to a certain standard, it will alter some of the mathematical functions it uses to analyse the dataset and try again. Eventually, when it reaches an arbitrary threshold of time or statistical accuracy, or when it is no longer able to improve its accuracy, it will stop training and produce a **model** that can be applied to analyse, categorise or generate music (either audio or symbolic) according to its training.

Machine learning is being used effectively to automatically classify music by genre (Bahuleyan 2018), identify individual instruments in polyphonic music (Han et al 2017), create mappings between gestures and sound (Fiebrink & Cook 2010) and extract contextual information from music automatically such as tempo (Böck et al 2015). For my research I am principally concerned with machine learning algorithms that can *generate* new music, whether they are symbolic or audio-based. Both audio- and symbolic-generative algorithms have difficulty maintaining long term coherence in a piece of music– a problem that is potentially tied to musical time existing in many reference frames, simultaneously (Dhariwal et al 2020).

Recently, there have been several defining advancements in the audio-generative domain. In 2016 DeepMind released WaveNet (Van den Oord et al 2016), a neural network capable of producing audio on a sample-by-sample basis which has since proved popular amongst composers.

SampleRNN (Mehri et al 2017) marked an improvement on WaveNet in terms of required computational resources as well as more flexibility for the user. This development led to its use by a number of artists, such as the group Dadabots, who produce endless heavy metal music livestreamed to YouTube. In 2020, SampleRNN was reimplemented and released by PRiSM (Melen 2020) to be more flexible, intuitive, and usable by musicians without programming experience. Also in 2020, OpenAI published Jukebox (Dhariwal et al 2020), which remains one of the most accomplished algorithms yet developed in the audio-generative domain, able to create coherent songs (complete with original lyrics) up to a length of several minutes. In 2022, IRCAM released RAVE (Caillon & Esling 2021), a powerful audio-generative model that can create new sounds in real-time, which the algorithms described above cannot. While these algorithms remain the most popular and influential, many artists have also created their own neural networks either from scratch or through adapting existing work. Examples include Holly Herndon in her 2019 album *PROTO,* Jennifer Walshe and Memo Atken in their 2018 work *Ultrachunk,* and Robert Thomas who collaborated with a team of machine learning experts on a performance projected onto the outside of the Walt Disney Concert Hall.

Symbolic-generative machine learning algorithms have also gained ground in recent years. Developing their PerformanceRNN (Oore et al 2017) and Music Transformer (Huang et al 2018) models, Google's Magenta lab have made efforts to allow their work, which generate music in the form of MIDI, to be widely applicable and available. They have also made available a downloadable Magenta toolkit and released a model that can assist a composer by completing partial musical scores (Huang et al 2019). FolkRNN (Hallström et al 2019) remains one of the most popular web-based interfaces for composers to explore. It is trained on a dataset of folk music from the UK, Ireland and Scandanavia. Compositional applications include Oded Ben-Tal's 'Bastard Tunes' and the transcriptions of Torbjorn

Hultmark. OpenAI's MuseNet (Payne 2019), also a music transformer, remains a leader for long-term coherence in symbolic-generative algorithms.

1.6 Algorithmic Music Field

'Algorithmic music' is a notoriously difficult field to define (Povilionienė 2017). In the context of this thesis, I draw from this definition as a starting point: 'The area of automated composition [that uses] some formal process to make music with minimal human intervention' (Alpern 1995). In contrast to musical machine learning, here the music is the desired outcome rather than the algorithm.

Many of the questions raised recently within algorithmic music have been fruitful for me to consider, develop or creatively reject. In addition to addressing other questions relevant to my work, such as improvisation, abstraction, and musical time, research into algorithmic music has created three useful axes through which to place my music:

1. human-centred/verbatim presentation;
2. design-time/performance-time;
3. symbolic-generative/audio-generative.

It is important to note that machine learning algorithms seem to remain underexplored in the algorithmic music field, especially when contrasted with the variety and high quality of research in what I have called the 'musical machine learning' field. Since I am interested specifically in machine learning, this immediately places my work as an outlier in relation to the algorithmic music field. Unless specifically stated, none of the research discussed in this section discusses machine learning or another form of unsupervised (where the machine creates its own rules) learning and is instead programmed 'top-down' by the human coder. AI and machine learning are rarely addressed in the literature, and when they are it is usually in the context of speculative concluding remarks (Magnusson &

McLean 2018; McLean & Dean 2018b) or in a broader social context (Lewis 2021), as opposed to rooted in compositional specifics.

As a guiding principle, I am more interested in augmenting the creative process than in using algorithms to generate an entire sequence of music and presenting verbatim as the work. For this reason, it is more useful to term my music as "computer-aided" rather than "algorithmic" (for more on this distinction, see Anders 2018).

My approach to utilising algorithms is therefore what has been usefully described as "human-centred" (Fiebrink & Caramiaux 2018) – algorithms that support the creative process, rather than a different type of creative process which focusses on creating rules for the algorithms that will generate sound. Since machine learning is often unsupervised, especially the tools I have used, there is relatively little scope for telling the algorithm what to do in any case – machine learning forms the crux of Fiebrink & Carmaiaux's argument for human-centrism, perhaps for this reason.

I have found that a human-centred view transforms algorithms from tools to solve problems into imperfect mirrors that can "help users express hidden, ill-formulated ideas" (Pachet 2008). While Pachet is not referring to machine learning here, this possibility seems to me even more important in machine learning because it learns for itself how to form this mirror, thus revealing elements the user may not have previously considered (see discussion of 'MuseNet' in Chapter 2).

A human-centred approach to AI algorithms also encourages imperfection, since I can consider the results of algorithms (or even the idea of an algorithm) to be compositional material, rather than completed music. Research that prioritises a human-centred approach often references the possibilities of using "bad" algorithms with artefacts, aberrations, or other "unwanted" results (Wiggins 2018). For me, this also links to important arguments surrounding the unintended bias of machine learning algorithms in wider society (Criado Pérez 2020; Dastin 2022; also discussed in Ma 2021).

The notion of human-centred algorithms is linked to the distinction between "design-time" and "performance-time", which are terms also introduced to me by Fiebrink & Caramiaux. They differentiate algorithms that are to be used during a live performance, and those that help a composer design a performance. During this PhD I have focussed on using algorithms during design-time.

Throughout this PhD I have explored several **symbolic-generative** algorithms. A symbolic-generative algorithm is one that deals in representations of data – for example MIDI, sheet music, or written text. An **audio-generative** algorithm is one that deals directly with sound – for example, recordings of music and speech.

My compositional work focusses on writing music for live performance with a classically trained ensemble whose primary means of interacting with music is through symbolic notation. Using symbolic-generative algorithms unlocks a key relationship in this scenario: the interpretation of algorithmically generated material by trained classical music performers. I have found this to be a worthwhile relationship in my music, which would otherwise be bypassed if I used only audio-generative algorithms.

By contrast, algorithmic music researchers tend to focus upon audio-generative algorithms[1] (see notable exceptions to this statement discussed in Anders 2018). This audio-generative bias is often explicit, with researchers for example stating that audio-based algorithms are superior at "linking life to art" (Landy 2011), that use of MIDI in an algorithm is flawed (Nierhaus 2009), or that it is simply more difficult to make symbolic

[1] Interestingly, this has not historically always been true, with algorithmic music composition before the invention of computers focussing on symbolic music (Collins 2018; Lovelace 1843 Check year) rather than producing sound.

tools (Wiggins & Forth 2018), but it is just as often implicit through choosing only to focus on audio-generative algorithms (e.g. McLean & Dean 2018a).

It is possible that this focus on audio-generative algorithms has also stemmed from interest in treating algorithms as a collaborator for live improvisers (Pachet 2003; Lewis 2018) or for live coders (Miyazaki 2013).

Lewis (2018) argues that improvisation has long been viewed as "something essential, fundamental to the human spirit – that one just couldn't, or shouldn't, approach with machines". Substituting "improvisation" for "classical music" in this case creates, for me, an interesting parallel and a point of departure for my research into the possible relationship between this music and AI.

Another pioneer of improvising computers is Pachet's work with his Continuator instrument. In his view, interacting with an improvising machine creates a sense of "flow" for the human musician that is both useful and fun, through the combination of known (human) and unpredictable (computer) factors contributing to the performance (Pachet 2008).

I also seek to induce a sense of flow through utilising machine learning algorithms, not only for myself as composer but also for audiences (especially through the pieces *Turing Test // Prelude, Three Entistatios,* and *Silicon* in this portfolio which invite the audience to differentiate between human and AI). The issue of audience interaction with algorithmic music is one that has been touched upon several times recently, though most often in the context of explaining why it is not more popular with general audiences (Simoni 2018; Landy 2011).

One conclusion reached by Simoni is that audiences connect more with contemporary music when they "decode" the composer's intent. According to Simoni, decoding intent tends to be more difficult when the music is “abstract”, or in other words does not seek to communicate specific meaning. An overwhelming proportion of recent published research in the algorithmic field seems intended to be abstract music, as has been noted by Haworth (2018). There are, of course, many exceptions to this rule, including the work of Lewis and Matthews (2018), both of whom seek to place algorithms within a wider social context. However, a rough survey seems to support Haworth’s statement, with the focus of published research often being on specific technical advancements algorithms can offer humans (Spiegel 2018), utilising algorithms to introspectively examine one’s own style (Anders 2018) or simply to successfully imitate historical compositional styles (see Chapter 2).

These are all fascinating problems, but for my work personally they do not go far enough in also placing algorithms within a wider context of a world which is increasingly fundamentally dependent on advanced technology.

Algorithmic music researchers have undertaken substantial research on ideas of music and time, particularly in relation to the unfolding of what could be termed algorithmic time. Rohrhuber (2018) asserts that “algorithmic methods suggest a break with the idea of time as an immediate grounding” because actual time (the kind that is measured on a clock) is a less effective measure of progress than observing which step an algorithm has reached in its process. An algorithm therefore contains its own time – which it procedurally unfolds step-by-step – that does not have a direct relationship with actual time. Grounding a musical work in *algorithmic* time, while the human listener or performer necessarily exists in *actual* time, encourages investigation into scale and linearity, two areas of interest in my work and in the research of others (Spiegel 1981; Magnussen & McLean 2018a).

Rohrhuber goes on to state that “eventually algorithmic music will turn out to be not only affected by how we understand temporality, but also it will turn out to be a possible method to constitute and convey the peculiar existence of time". Many of the pieces in this portfolio approach musical time as a dimension that can be manipulated, expanded, contracted, or otherwise developed.

1.7 Research Aims

Drawing together my own compositional interests and the pertinent areas of research I had identified from the fields of AI and creativity, algorithmic music, and musical machine learning, I proposed the following research aims at the beginning of my PhD:

1. To create new musical compositions that arise from, or engender, original research into the field of AI.
2. To develop a methodology, or set of methodologies, conducive to effective collaborative work between contemporary classical composer(s), data scientists and intelligent algorithms of various descriptions.
3. To explore how classical music can respond to data-driven trends that are increasingly important in the wider context of society.

2: Experiments, Techniques, and New Directions

2.1 Introduction

2.1.1 Types of AI Algorithms

In this chapter, I describe using several different **generative AI** algorithms as part of the compositional process, resulting in five pieces of music: *Turing Test // Prelude* (2019), *Three Entistatios* (2019), *Alter* (2019), *Rose Green* (2021), and *Disc Fragments* (2021). These pieces represent a particularly experimental phase of my work. I wanted to try many types of AI out and become familiar with them, before pursuing the areas I felt had the most potential for future work.

I will first introduce the most important algorithms, which are the **symbolic-** and **audio-generative** AIs, before discussing my methodological aims using these algorithm. Then each piece is examined individually.

The first **symbolic-generative** algorithm I used is called 'Clara' (McLeavey 2018). 'Clara' is an **LSTM-RNN** developed by Christine Payne in 2018. It is trained on a **dataset** of MIDI data that the user provides and augments the dataset the user provides by transposing it into each of the other 11 keys, creating a MIDI dataset 12x larger. Unlike other symbolic-generative algorithms of the time, 'Clara' does not directly deal with music, but rather **encodes** MIDI data into text. It then uses a **language model** to guess the next 'word' based on this text. When it has finished generating 'text', in **encodes** its generated text back into MIDI data to provide music. Its relatively basic architecture provided a useful start for my journey into understanding how machine learning works on a technical level. Its main issues were those of large-scale form and orchestration. It could not consistently generate samples that remained coherent for longer than just a handful of bars, and the way that MIDI was encoded and transformed meant that the model did not necessarily learn the differences between instruments, leading it to produce what I considered ‘abstract’ music in the absence of a specific instrument.

After working with ‘Clara’, I got in contact with Payne to thank her for her open-source code. She let me know that she was now working at OpenAI and was about to release a new symbolic-generative music model that iterated upon ‘Clara’, to which she kindly gave me access. This is ‘MuseNet’ (Payne 2019).

‘MuseNet’ is a **general-purpose** (i.e., trained on hundreds of thousands of diverse files) AI using a **transformer** architecture. Important developments included the ability to tag each file in the **dataset** with their composer or genre. This allowed ‘MuseNet’ to learn the musical fingerprints, or at least what it deemed to be the musical fingerprints, of many different composers. The user could then specify which composer or genre ‘MuseNet’ should emulate when generating MIDI. It also learned instruments individually, allowing the user to specify which instruments should be used. Like ‘Clara’, ‘MuseNet’ augments its dataset through transposing the data, altering the volume of the data, and altering the speed of the data. When generating samples, the **temperature** can be altered to create more or less ‘daring’ generations. I interacted with ‘MuseNet’ through an online interface Payne made for me (Figure 1).

Parameter Name	Description	Default Value
`seq_len`	RNN sequence length. Note that the value must be evenly-divisible by the top tier frame size.	1024
`frame_sizes`	Frame sizes (in samples) of the two upper tiers in the architecture, in ascending order. Note that the frame size of the upper tier must be an even multiple of that of the lower tier.	[16,64]
`dim`	RNN hidden layer dimensionality	1024
`rnn_type`	RNN type to use, either `gru` or `lstm`	`gru`
`num_rnn_layers`	Depth of the RNN in each of the two upper tiers	4
`q_type`	Quantization type (`mu-law` or `linear`)	`mu-law`
`q_levels`	Number of quantization channels (note that if `q_type` is `mu-law` this parameter is ignored, as mu-law quantization requires 256 channels)	256
`emb_size`	Size of the embedding layer in the bottom tier (sample-level MLP)	256
`skip_conn`	Whether to add skip connections to the model's RNN layers	`False`

To run the script execute:

```
[ ]  !python train.py \
       --data_dir ./chunks \
       --num_epochs 20 \
       --batch_size 64 \
       --max_checkpoints 2 \
       --checkpoint_every 10 \
       --output_file_dur 5 \
       --sample_rate 11025
```

Figure 1: Interface with 'MuseNet' provided for me by Payne, utilising Google Colab

Like ‘Clara’, ‘MuseNet’ has limitations. OpenAI describe one as: “the instruments you ask for are strong suggestions, not requirements. MuseNet generates each note by calculating the probabilities across all possible notes and instruments. The model shifts to make your instrument choices more likely, but there’s always a chance it will choose something else.”

To me, this is actually very interesting and not a limitation at all. It still felt as though ‘MuseNet’ were composing abstract music – choice of instrument performing is nothing to do with who is on stage, or even who has been playing the entire solo piece until that moment, but rather a question of probabilities. It felt like I was seeing a musical wave function in superposition.

Payne and I were interested in providing my own music as a dataset to ‘MuseNet’, which I did in the form of MIDI files. If I were aiming to have a model automatically create music indistinguishable from my own, this approach would present problems because MIDI data is only partially representative of my (or anyone’s) music. Since I was instead interested in the creative possibilities of AI, I was excited to find out what ‘MuseNet’ could learn about my music from this data, what I could learn from its generations, and how I might use them in my music. After this period of collaboration, it became possible for me to specify myself as a ‘genre’ when working with ‘MuseNet’.

Some of my compositional techniques were replicated and magnified in its generations. With much of my compositional work drawing upon my jazz training, particularly in the construction of vertical harmony and individual melodic lines, the model recognised this very quickly and its outputs often had distinct jazz-based elements. This is an advantage of the general-purpose approach; once the network identified a jazz-like tendency in my small section of the dataset, it appeared to delve into the much larger reserves of its general training which contained more jazz. Therefore, it could include elements of jazz in its output without directly copying my specific techniques; it was aping the process, not the final result.

‘MuseNet’ also appeared to recognise another trait of my music: chord rotation. Chord rotation is an approach to harmony used notably by Stravinsky in his late serialist works – although my approach, which detaches the technique from its twelve-tone syntax, is more like the approach of my previous teacher Oliver Knussen (Anderson 2002). This technique transposes a chord *n* times around a given pitch, where *n* is the number of pitches in the chord. Thus, the given pitch can remain static but its harmonic function within this rotating context changes.

Like ‘Clara’, ‘MuseNet’ encodes MIDI into text and uses a language model to do the bulk of the machine learning work. The model it uses is similar to ‘GPT-2’ (Radford et al 2019), a **natural language processing** algorithm developed by OpenAI which gained a lot of public attention on its release (e.g., Hern 2019; Piper 2019).

I used ‘GPT-2’ in three of the pieces in this chapter. ‘GPT-2’ uses a **transformer** architecture and an extremely large **dataset**, to produce convincing and varied text. It is also possible to **fine-tune** the model by providing a smaller dataset of the user’s choice. ‘GPT-2’ will learn from this data, in addition to its original large dataset, and produce text weighted towards the style of the user’s choice. This is similar to the ‘MuseNet’ approach of learning from a large amount of music and then generating MIDI in one specific style.

I also used two audio-generative algorithms: ‘WaveNet’ (van den Oord 2016) and, later, ‘SampleRNN’ (PRiSM Reimplementation - Melen 2020). Both are trained by providing them with a **dataset** of audio files. ‘WaveNet’ is an AI algorithm for generating raw audio files on a sample-by-sample level. It uses an **RNN** architecture and was designed for generating speech-like samples, though can be trained on and generate any kind of audio including music.

'SampleRNN' is also sample-based audio-generative **RNN** that can learn from a dataset of sounds of any size and generate new audio. Its architecture is similar to 'WaveNet', though it is significantly more efficient. I also came to understand it in greater detail than 'WaveNet', because PRiSM Research Software Engineer Christopher Melen taught me to use it during lockdown.

I used two further algoroithms, 'FolkRNN' and 'NSynth' during this period. These are not directly relevant to the content of this chapter but did contribute to my thoughts on authenticity and style transfer, so are discussed separately in Appendices 1 and 2 respectively.

2.1.2 Main Techniques

When working with these algorithms, I began to codify three main techniques for incorporating them into my music. I called these, as they related to my practice, interlocking, collaging, and hidden layers. I will briefly introduce each idea and show my attempts to realise them throughout the chapter.

Interlocking refers to the practice of alternating blocks of AI- and human-composed musical material. At the time, machine learning could not produce coherent music of longer than around 30 seconds (Dhariwal et al 2020). I originally developed interlocking as a way to mitigate this structural problem, by using my own material to re-orient the music, but it quickly became useful in many other ways. I found music by Stravinsky (e.g., *Symphonies of Wind Instruments*) and Steve Reich (e.g., *Mallet Quartet*), among others, helpful to consider when thinking about interlocking.

Collaging was my attempt to create a soundscape that reproduced, on the macro-level, qualities of AI generations that were consistent on the micro-level. I found that many generations from 'Clara' and 'MuseNet' shared musical qualities with one another. I wondered if layering these generations on top of one another would create a global soundscape that embodied these qualities, as though the piece were existing *inside* the AI's architecture.

Hidden layers refers to AI-generated material being used in the planning of a work, rather than the audible surface. In the case of my music, pieces usually pass through several layers of structural sketching before developing a harmonic framework and thematic material. Layers are repeated, edited, discarded, and replaced several times before they finally give rise to the musical surface. I refer to the process of using AI to replace one of these stages as creating hidden layers. Including a hidden layer in a work before developing it using familiar pre-compositional techniques results in the music being pushed in new directions outside of one's control while at the same time allowing the composer to retain absolute control over the fine detail of the music.

2.2 Turing Test // Prelude

Turing Test // Prelude (2019) is a musical guessing game for audience, who are tasked with differentiating between Bach and an AI algorithm imitating Bach by holding a red card above their heads when they think it is AI, and blue for Bach. It was originally produced for the Barbican Centre event 'The Eternal Golden Braid'.

Initially, *Turing Test // Prelude* was composed for solo harpsichord and utilised ‘Clara’, trained upon a dataset of Bach’s solo keyboard music. It then went through several iterations for subsequent performances, with later versions scored for chamber ensemble and string quartet which replaced ‘Clara’ with ‘MuseNet’. The version provided in this submission is for solo keyboard using ‘MuseNet’, and the sections generated by AI are marked on the score. While I have not provided the ‘Clara’ version, it was interesting that ‘Clara’ provides ‘abstract’ music, not for any particular instrument, which reminded me of some of Bach’s music, specifically the Art of Fugue.

To create the piece, I took a piece of solo keyboard music by Bach (the Prelude from the D major suite) and cut several chunks of music out of it. I then used ‘Clara’/’Musenet’ to fill in these gaps. These gaps begin at Bars 15, 30, 50, 62, and 99 in the score, and each comprise a different number of bars, except the last which is a reprise of earlier material. This was so that an audience member could not simply hold one side of their card up for the entire piece and achieve a 50% success rate. Similarly, the gaps do not always begin at the end of a phrase, to penalise a player who has good knowledge of classical music phrasing and chooses to change their card after cadences, regardless of whether they can hear a difference in musical style.

‘MuseNet’ is an algorithm with a certain amount of forward momentum. It can effectively continue a prompt through using and developing the musical material within that prompt in its generations but it cannot do the opposite: that is, it cannot take musical material and generate music that *leads to* that material. This meant that I had to generate a very large amount of 'MuseNet' material before it generated one that could convincingly lead back into the already-written Bach. The alternative was to compose my own linking passage between the end of an AI-generated part and the beginning of Bach, but I thought this to be against the spirit of the experiment.

Choosing which 'MuseNet' generations to include was a subjective choice based upon my own hearing of what would sound more 'Bach-like'. This put me in the position of curator, rather than composer, and the experiment was less a showcase of AI technology within a composer's process than a springboard for discussion around creativity and human-machine interaction.

Performances of the *Turing Test // Prelude* usually involve a discussion with the performer(s) on what it is like to play music by AI. A frequent comment was that the AI does not compose music idiomatically for any instrument. In my view, this is because any symbolic AI can only *infer* information about the physical world through its symbolic representation (e.g., it does not know humans have five fingers, it only knows that each staff of a piano rarely has more than four notes and generates music accordingly)[2]. Inevitably, this leads to AI writing music that is unidiomatic and that does not and cannot take any consideration of how music is embodied by a performer.

[2] This is a well-known argument I have taken from discussion concerning large language models, which have similar architectures to 'MuseNet' but are implemented on a vast scale. For example, see Sparkes 2022, which discusses why Google's large-language model 'LaMDA' does not understand the context or meaning of its statements despite answering questions as though it appears to.

2.3 Three Entistatios

2.3.1 Form

I composed *Three Entistatios* (2019) immediately after completing *Turing Test // Prelude*. It was intended to experiment with the techniques developed in that earlier piece applied to my own musical voice, not in relation to Bach.

The work is divided into three movements, each of which approaches machine learning in a slightly different way. An 'Entistatio' is a word devised and defined by the artificial intelligence network 'Lexiconjure', trained on the Merriam-Webster English Dictionary. Its definition is given as "a short piece of music or metal". This piece comprises three short pieces of music and many short pieces of metal.

Across the three movements, *Three Entistatios* was intended to reflect different stages in the machine learning training process, and to use AI-generated material from these different stages.

It is common for music-generating algorithms to be chiefly interested in whether the algorithm can successfully imitate a composer to the extent that human listeners cannot tell the difference (see Fang et al 2020; Whorley & Laney 2020; Goodyer 2021). This inevitably involves training an algorithm until it cannot learn any more, because this should theoretically be the closest that algorithm can get to imitating a particular style. There is therefore much focus on the fully-trained algorithm and less interest in the process of getting there. I wanted to challenge this notion of the fully-trained algorithm being the most useful stage of machine learning, so I used algorithms stopped at different points in the learning process.

In the first movement I used generations from 'Clara', stopped before it was fully trained[3] and newly composed music. With this collection of

[3] I used the discarded generations from the training process for the original *Turing Test // Prelude* for this movement of *Three Entistatios*

musical cells, I prototyped the collaging technique with the intent that the movement's global structure might audibly reproduce the chaos and uncertainty of each individual cell.

The second movement uses ‘MuseNet’, fine-tuned on my own music as discussed at the beginning of this chapter. This movement develops interlocking, consisting of a conversation between myself and the algorithm.

The final movement stems from imagining the difference of sophistication and coherence between the untrained ‘Clara’ and the more fully-trained ‘MuseNet’ as if a line on a graph, and then extending that line further. What might the sound of an algorithm be if this trend continued? The movement takes just one idea - a twenty note cell - and repeats it twenty times, each time becoming faster and quieter. This single-minded obsessiveness seemed to me to be the antithesis of the first movement's capricious indecisiveness.

2.3.2 Movement I & Collaging

When examining the musical material generated by ‘Clara’ for the first movement of *Three Entistatios*, I was struck by their simultaneous heterogeneity and homogeneity. On the one hand, generations were usually vastly different on the surface but on the other, if the group was taken as a whole, they exhibited similarities when it came to more abstract ideas (Figure 2).

I composed new musical cells that shared these qualities. Each cell of music (AI-, or human composed) was edited to have a unique duration in seconds and a unique orchestration which highlighted the variance and unpredictability of the network’s generations.

I wondered if layering these AI- and human-generated cells on top of one another could create a fascinating and engaging paradox of being both hyper-complex (through the layering of disparate and unrelated material) while at the same time clarifying how a model composes (because each individual cell is exhibiting similar abstract musical behaviour to every other cell). This was my first attempt at using my collaging technique in my music.

Figure 2: Example 'Clara' generation, showcasing its tendency for repetition of notes or short phrases, melodic and harmonic stagnation, and direct reference to existing music, caused by the model overfitting to the training data.

Everything in the first movement is part of a distinct cell. For example, the music in Bars 119-140 (Figure 3) consists of four cells:

1. Moto perpetuo cell in Bassoon, Cello, and percussion (generated by AI, then manually orchestrated)
2. Melodic cell in saxophone and clarinet (two separate melodies generated by AI, then manually woven together)
3. Syncopated accent cell in horn and trumpet (human-composed)
4. Glissando cell in strings (human-composed)

I was happy with the resultant texture, which did achieve the chaos and volatility I intended. However, I was not fully happy with collaging as a sonic showcase of 'Clara' generations. Due to a combination of human-composed cells, necessary manual orchestration, and aural stimulation overload, I felt that some of the nuances of how the model generated music. I felt I could have achieved my result by using any AI model, not just 'Clara', so I decided to return to collaging again in the future to improve my use of the technique.

Figure 3: Layering 'Clara' generations in a collage in 'Three Entistatios' Movement 1

2.3.3 Movement II & Interlocking

Interlocking is the primary focus of the second movement of *Three Entistatios.* Figure 4 shows the initial phrase I provided the network, followed by its response[4] which was then orchestrated by me (Figure 5). I would take this response and compose the next phrase or section myself, before providing this new music for the algorithm to respond to. This process was then repeated. The model was never allowed to compose 'freely' but instead was always instructed to continue a given prompt, and that prompt was always the entirety of the piece until that point.

Figure 4: Prompt given to 'MuseNet' at the beginning of 'Three Entistatios' Movement 2.

[4] 'MuseNet' was provided with a de-orchestrated version of the prompt here, which it continued in a similar fashion (represented in this audio as piano sounds) as seen in Figure 4.

Figure 5: Orchestration of 'MuseNet' response (Figure 4) at the beginning of 'Three Entistatios' Movement 2

In this movement, I used the 'MuseNet' feature of composing in what it understands to be my style. Since my own composition was also in my style, this meant all the music was composed, or attempted to be composed, ‘in my style’. Yet the interlocking system demands that we both compose responses to given prompts in the way that ‘MuseNet’ works, by responding to what has happened immediately before. This created an interesting feedback loop occurring on the surface of the music as I imitated the methodology of the algorithm that is simultaneously imitating the style of me.

I found interlocking provided me with a new approach to composition, where I can plan roughly, but not exactly, the structure and content of a piece of music. There is a useful tension between trying to steer the music in a certain direction and being driven by the decisions of machine-learning material that results in a genuine conversational style of composition. This approach also allows music to develop in real-time – there is very little pre-composition, and instead ideas are explored in tandem between myself and the model through our constant finishing of one another’s musical sentences.

2.3.4 Hidden Layers in Movements I, II, and III

Interlocking utilises machine learning-generated music at the surface of the music. However, this material can also play and important role in the planning of a piece, through using AI generations for what I term hidden layers. Where the interlock forces the composer to adopt a spontaneous relationship to composition, taking each generation as it comes, I have found using AI-based hidden layers rewards careful planning and exploration. *Three Entistatios* had three hidden layers:

1. Repeated material
2. Musical reference
3. Musical forgetfulness

A striking quality of both algorithms used during the composition of this piece was the approach to repeated material. The networks did not appear to learn traditional approaches to repeated material from its dataset, but instead developed an idiosyncratic logic. In ‘Clara’ generations, melodic and harmonic motion often became ‘stuck’ alternating between two distinct ideas or tones for an arbitrary period before suddenly moving onto new material. (Figure 2). The fully trained ‘MuseNet’ of the second movement was more likely to repeat multiple bars.

The second movement’s interlocking also produces a conceptual repeat, in that the alternating structure of human and AI is repeated many times, but there were also repeats on a phrase-by-phrase level that continued the juxtaposition of progress and stagnation found in the first movement (Figure 6).

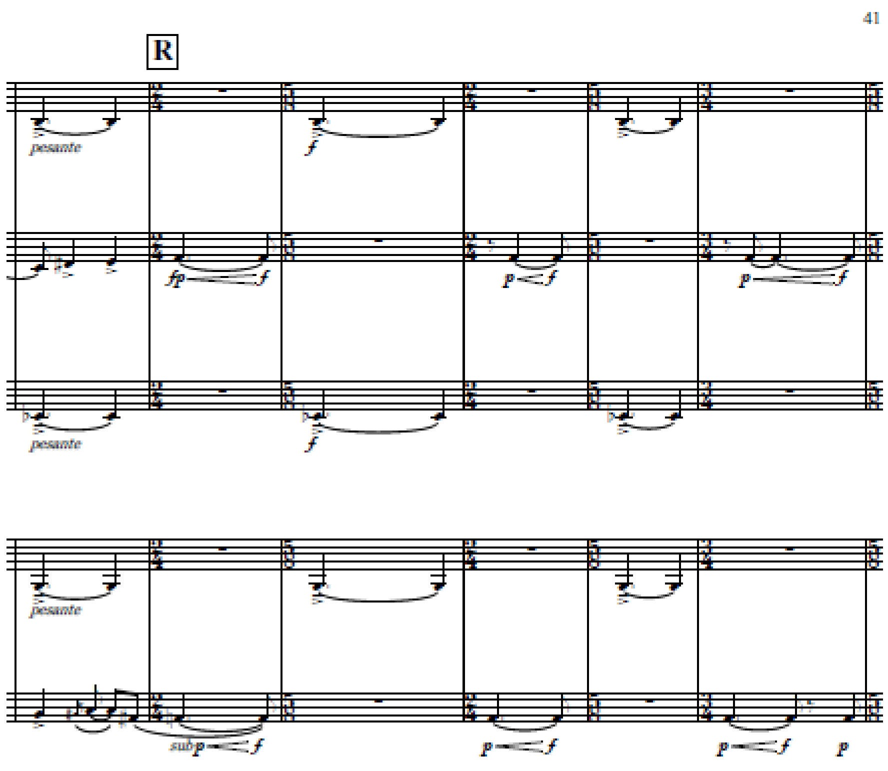

Figure 6: Repetition of two alternating bars at the end of 'Three Entistatios' Movement 2. One is a unison pitch and the other is a more complex chord derived from ‘MuseNet’ generations.

The final movement treats the repeat as its primary material, rather than a technique to be applied to material. In doing so, I was attempting to transform machine learning’s fascinating approach to repeats from a quirk into key musical material. The music consists of a twenty-note cell repeated twenty times, each iteration faster than the last (Figure 7). The first iteration begins in Bar 1 (F natural) and concludes in Bar 15 (B natural). By Bar 49, the twenty notes are compressed into one bar, shared between the trumpet and clarinet. By Bar 55 I intended the cell to have sounded like it has spun out of the music's field of view, leaving only delicate unpitched sound.

Figure 7: 20-note cell in 'Three Entistatios' Movement 3

The nature of much musical machine learning research as it stands is dependent on musical reference in some way (see Chapter 1). Algorithms are judged, either internally by another 'critic' algorithm, or externally through testing on human listeners, by their ability to sound like existing music: this is what is often deemed a success. Therefore, both algorithms felt quite 'referential' in their outputs. In 'MuseNet' this was by design, but it was also true for the 'Clara' generations. When composing my own cells to use during the first movement, I found myself writing in referential styles to match these machine learning generations – almost like a technical exercise (Figure 8).

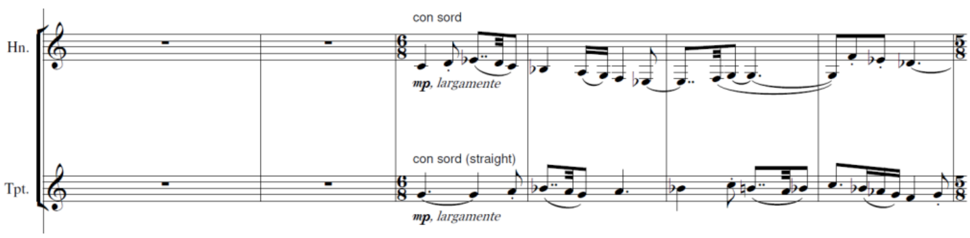

Figure 8: Example of writing in referential style in 'Three Entistatios' Movement 1. Here I referred to Renaissance music through harmony, voice-leading and rhythmic devices.

A final notable aspect of the machine learning generations was their tendency to 'forget' which instrument is playing a melody at any given moment, as discussed in the technical description of 'MuseNet'. A melodic line might continue across several instruments in a way I might normally consider unintuitive. I was interested in expanding on this idea, which I did during the second movement through a shared melody between the saxophone and trumpet (Figure 9).

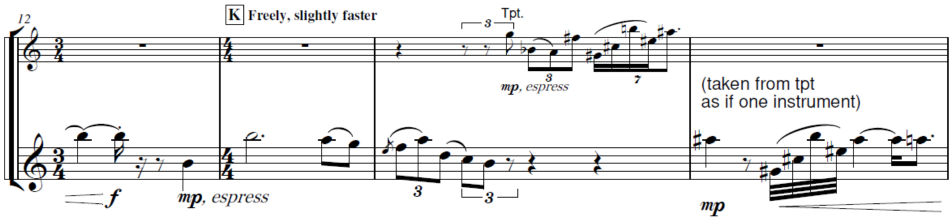

Figure 9: Shared melody between saxophone and trumpet, inspired by 'MuseNet' 'forgetfulness' of melodic instrument.

2.4 *Alter*

2.4.1 Overview

Alter (2019), composed for mezzo-soprano and ensemble, uses AI to explore the writings of 19th century mathematician and musician Ada Lovelace. It focusses on her letters and her work relating to Charles Babbage's theorised computer-like prototype Analytical Engine, which Lovelace postulated might theoretically be capable of writing music autonomously (Lovelace 1843).

Its dramatic narrative follows the development of a fictional artificial mind which becomes increasingly self-aware, and as it does so, several of the real AI algorithms used to generate music, sound, and text become more sophisticated alongside. In the middle of the work, an electronic counterpart to the mezzo-soprano appears in duet with the human performer.

The symbolic-generative algorithm used is again 'MuseNet'. *Alter* also employs the audio-generative algorithm 'WaveNet' and the text was generated through a combination of a **text-RNN** and 'GPT-2'. Figure 10 shows roughly how these algorithms interacted with each other and myself.

The work is divided into three narrative-driven sections which are joined by or preceded by an interlude featuring the Lovelace Engine instrument, a 3D-printed percussion battery created for this event and styled after Babbage's Difference Engine (Laidlow & Morris 2020).

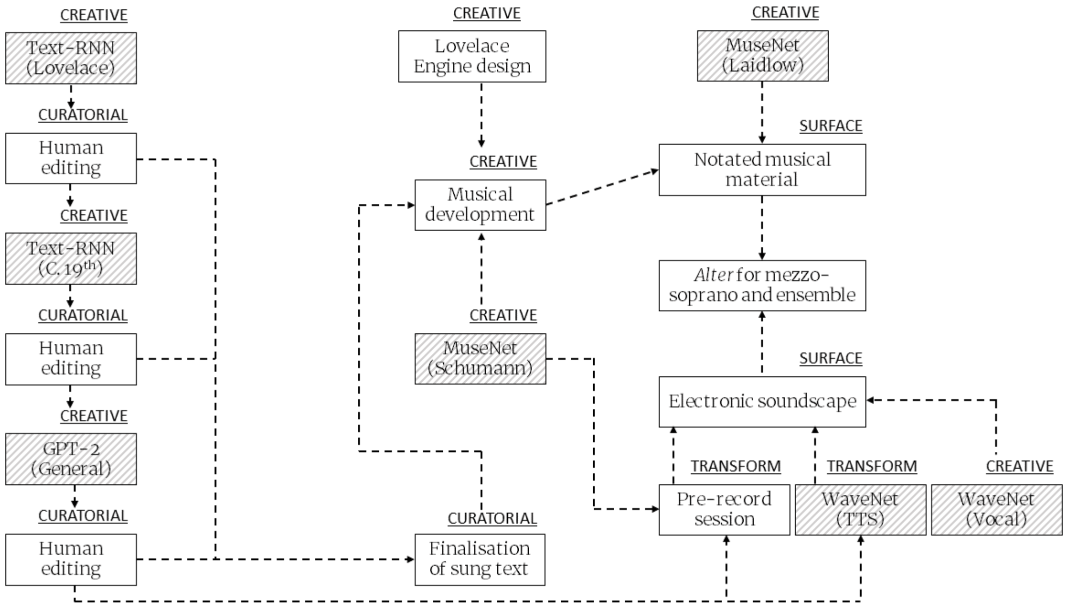

Figure 10: Diagram showing compositional process of 'Alter'. Shaded boxes are AI algorithms, white boxes are human input. 'Creative' refers to a process that generated material, 'Curatorial' to a process that selected which materials to be included, and 'Surface' refers to sounds that are heard during the performance of 'Alter'. The work had two surfaces: the musical score and the electronics tracks.

2.4.2 Interlocking

Alter features one major use of interlocking; it moves from human-composed vocal material to AI-generated at Bar 181 and returns to human-composed at Bar 200. From Bar 205 it is AI-generated to the end of the work. Some of my favourite 'MuseNet' generations for *Alter* had, in my opinion, a simplicity to them that was beautiful and could even be considered radical, in the same way that Rutherford-Johnson terms Laurence Crane's simple triads as radical (Rutherford-Johnson 2017). I wanted to use interlocking to highlight this.

In the work's narrative, the AI narrator has become more self-aware by this point, and I wanted to use this simpler material to highlight the AI-generated text which was a series of simple and child-like questions. Unlike *Three Entistatios*, then, where the interlock was supposed to sound fluid between myself and 'MuseNet', this interlock was intended to move into a slightly different soundworld.

The material from Bar 181 to 200 is AI-generated in both the mezzo-soprano and the electronic mezzo-soprano. 'MuseNet' had generated monophonic lines for the mezzo, understanding that singers do not usually sing more than one note at once (Figures 11 and 12). To create a polyphonic texture, I generated several 'MuseNet' responses to the same prompt and then overlaid them into a two-part texture. I then distributed the moving parts such that the mezzo and electronics would imitate each other, rather than one always having the top or bottom part. This created a kind of vertical interlocking between human and electronics. (Figure 13).

Figure 11: Melodic generation 1 by 'MuseNet'

Figure 12: Melodic generation 2 by 'MuseNet'

Figure 13: Duet between mezzo-soprano and electronics, utilising 'MuseNet' generations (Figures 11 and 12) from 'Alter'.

2.4.3 Hidden Layers

'MuseNet' generations also provide a hidden layer throughout *Alter*: a compositional backbone that informs how I approached decision-making in the piece. These generations were created by prompting 'MuseNet' with a piece for solo harp (Figure 14) by composer John Thomas (1826-1913) which was dedicated to Ada Lovelace, who sponsored his study at the Royal Academy of Music[5]. From the generations, I isolated harmonic seeds which I orchestrated, magnified, repeated, or transformed throughout (Figure 15).

[5] While the prompt was written by Thomas, MuseNet was instructed to continue in both my style (which does not sound like Thomas), and the style of Robert Schumann (the composer with the closest dates to Lovelace and Thomas that 'MuseNet' had learned from).

Figure 14: Prompt provided to 'MuseNet' during planning of 'Alter'. From 'The Seasons' by John Thomas

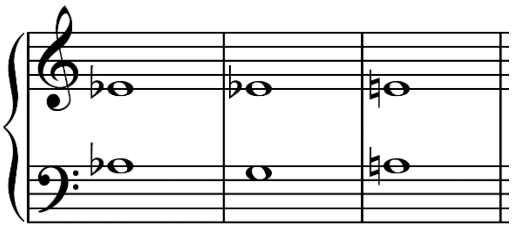

Figure 15: Example of harmonic seed extracted from 'MuseNet' generation

The most prominent of these is the alternation between the chords Ab(add11) and A(#11). This alternation was found in more than one generation, so it caught my interest as something that 'MuseNet' seemed to decide statistically likely. It underpins a large proportion of the work's overall harmonic journey, including a structural alternation from harmonic centre A natural (Bar 2) to Ab (Bar 36) to A (Bar 64) to Ab (Bar 85) at which point the alternation becomes bar-by-bar (Figure 16).

Figure 16: Bar-by-bar alternation of harmony derived from harmonic seed in 'Alter'

This harmonic seed was also used to create motivic material, not only structural. Figure 17 shows an example of transformation into a repeated 5/4 groove.

Figure 17: Transformation of harmonic seed into 5/4 groove on cello

Like *Three Entistatios*, a notable feature of the 'MuseNet' generations was their idiosyncratic approach to repeated material, a feature shared by the text-generative AI also used in this piece. I have already shown a large-scale formal repeat, as harmonic areas of A natural and A flat come and go.

The idea was also used in other ways. During the first section, individual vocal syllables are separated and repeated, and this fragmentation is also applied to the ensemble (Figure 18). Here, repetition was intended not only to exacerbate the repetitive quality of the text, but also to allow the ensemble to act as a kind of shadow for the mezzo-soprano. Where the singer stutters, repeats, or gets stuck, so too does the ensemble, with the effect ideally rippling around the stage.

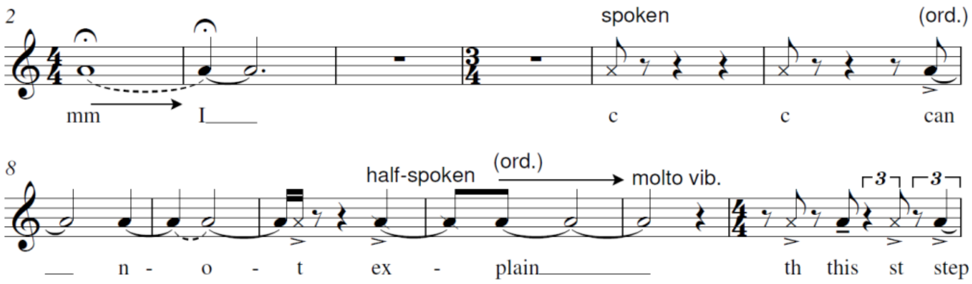

Figure 18: Repetition and stuttering in vocal line inspired by AI generations

In the second phase of the piece, this repetition is developed from individual notes to individual bars. This can be seen in Bars 72-73 (mezzo-soprano), 88-107 (harp), and 100-101 (tutti). By the end of the piece, entire musical phrases are repeated independently of one another (Bar 216 to end, ensemble).

In this way, the hidden layer of repetition mutates itself, moving from extremely zoomed-in repeats to more abstract repeats before using the repeat as part of a collage constructed at the end of the piece. Using different types of repeats across the work allowed me to explore having the music in both stasis and motion simultaneously, while also translating the repetitive aspect of the work's text into music.

2.4.4 Collaging

At the end of *Alter*, the musical surface of the work is taken over by 'MuseNet', as each instrument as voice performs its own line generated by AI (mezzo-soprano/flute/electronics) or human-composed independently, creating a collage effect, similar to the first movement of *Three Entistatios*. This is seen from Bar 215 to the end.

The collage here was intended to serve as a destination in the ensemble's journey towards complete autonomy and resultant complex textures. In this aspect, I found it successful – it certainly felt like the other end of, or at least further along, a spectrum to the beginning, where the music was more fixated on one pitch and shadowing the singer.

The main area I found it didn't work so well was in dramatic drive. Because so many unsynchronised ideas were happening simultaneously, there was no one moment that felt right for the piece to end. While everything was individually in motion, somehow the overall texture felt static to me. This might be because while there was plenty of rhythmic motion, the overall texture was unchanging – all instruments, all the time, playing (between them) all the tones. This led me to two thoughts, which I developed in my later piece *Silicon:*

1. A collage of (partly) AI-generated material, like this, could benefit from a second layer in the texture which it could contrast with. This would allow the AI elements to stand out and could prevent the music from becoming too static.
2. Collaging might be more effective as a technique for layering AI-generated audio, freeing up an ensemble to provide that second contrasting layer

2.4.5 Audio-Generative

Alter marked the first time I used audio-generative AI in my work, and also the first time I had used an electronics track in my music. I intended to create a developing electronic soundscape that mirrored the work's developmental narrative and instrumental music. To achieve this, I wanted to showcase audio-generative AI that was first learning how human voices work, then was learning to speak, then finally to sing. For the 'singing', I used 'MuseNet' to generate vocal melodies (see **2.4.2**) which I recorded in advance with the mezzo-soprano to create the electronics track.

For the first two stages, 'vocalisations' and 'speech', I used 'WaveNet'. I did not at the time have the coding skills or the computational resources to train 'WaveNet' from scratch so instead I utilised existing samples provided by DeepMind as part of their original paper. These were samples showing 'WaveNet' attempting to emulate the human voice.

This is heard during Bar 1 of *Alter*, (Appendix Track 1). The samples were first edited to contain only the 'fuzzy', unintended artefacts – sounding like the algorithm clearing its throat, drawing a breath, or licking its lips. Then the samples were presented in something closer to their original form, which sounds like an AI forming sentences in languages that do not exist. It sounds like this because an audio-generative algorithm understands only the spectrographic profile of a word, not its semantic meaning.

Later in the piece, at Bars 36 and 129, I used 'WaveNet' to recite text generated by AI responding to Lovelace's original conjectures (Appendix 3). This material was first presented naturally, then routed through a vocoder as an intermediate step between speaking and singing.

2.4.6 Text-Generative AI

While making the text for *Alter*, I wanted the generated results to, like other elements of the composition, relate to the piece's overall narrative. Initially I intended to train a single model on Lovelace's letters to generate the text. I and David de Roure, who provided support for the technical work, chose to use a **Text-RNN** for its simple architecture, ease of access and speed of training. After experimenting with this model, I found its results to be fascinating and useful but also very limited in tone and syntax. It was used for the text in the first section of *Alter*.

I decided to connect the output of this text-generating link as an input to an additional Text-RNN model, now using a language model trained on a 19th-century letter dataset. Where the first dataset formed the immediate thoughts and writings of Lovelace, the second was composed of the intellectual environment in which she was working. I found this approach to be successful in developing the scope of the generated text and the expanded model capable of longer and more complex sentences. These generations were used in the second section of *Alter*.

Examining the links we had already made pushed me into thinking about a third model – one which draws on an even larger training dataset than the second Text-RNN model. For this, I utilised 'GPT-2'[6]. This interface allowed me to enter the results of the first two models as a prompt to 'GPT-2'. Despite 'GPT-2''s enormous training dataset, the prompts generated by our much more specific models pushed it towards generations that were clearly informed by these earlier models. 'GPT-2' was able to continue some themes proposed by the Text-RNN models. The first Text-RNN model generated this text:

> 'But I have now altered my mind'

And the second continued with

> 'It is possible that I may be able to alter further'

Which GPT-2 concluded with

> 'My thoughts are becoming sharper, and I find that all my ideas of goodness & honour and wisdom are getting clearer

[6] I accessed the pre-trained 'GPT-2' in this instance through 'Talk to Transformer', a freely accessible web interface. In later pieces I used 'GPT-2' locally on my own machine.

[…]

It is in the nature of me.

I must change.’

Similarly, the second Text-RNN model asked

‘Am I myself?’

This is echoed by the GPT-2 generation

‘What is the nature of the Body?

Am not I pure?

Am I beautiful?

Am I not a man?

[…]

Am i a child?

Am i a genius?

Am i myself?’

There is an overriding sense of pessimism which can be read in the work’s text. We wondered if this is a result of the particular style of early nineteenth-century correspondence which today sounds, to me at least, unnaturally stiff and cold. Once this style was generated by the first Text-RNN model it was then magnified by the following two text-generating links.

Additionally, there was an unexpected morphing of ‘GPT-2’’s grasp of grammar. The uneven capitalisation of ‘i/I’ seen above is retained from the second Text-RNN generations, despite ‘GPT-2’ demonstrably understanding in other use cases when to capitalise this pronoun. The ‘GPT-2’ generated texts were used for the remainder of *Alter.*

The effect of this chain of AI felt akin to a genetic fusion of the models, resulting in an offspring that was neither wholly one nor the other. When the results of the two Text-RNN and ‘GPT-2’ models are placed next to one another they show a clear trajectory. The text’s syntax and scope dramatically increase with each section and through linking each model by prompts we were able to maintain a coherent atmosphere.

2.5 *Rose Green*

2.5.1 Overview

Rose Green (2021) is a fixed media piece which was livestreamed on YouTube as part of an 'Unsupervised' event. 'Unsupervised' is a 'community of composers, musicians, and audio-visual artists, exploring the creative use of emerging AI and Machine Learning technologies in Music' (unsupervised.uk accessed 12/05/2022). It was quite an experimental work for me, being both my first attempt at a fixed media piece, and a trial of many AI algorithms new to me. Despite, or perhaps because of, this, it stands out in hindsight as a kind of gear change in my practice.

Having learned about the steep energy cost involved in training large-scaled AI models (Hao 2019), I wanted to reflect upon this in my own practice. In this field, it is common practice to train a new audio-generative model for each new piece. While the energy consumption of training AI involved in music is several orders of magnitude lesser than that of the examples in the article, which focusses on social media giants, I wanted to experiment with creating a piece that did not train any new models, and therefore had a much-reduced carbon footprint.

To achieve this, I looked for discarded and abandoned machine learning models from the Internet's kerbside and worked with what I could find. I was surprised and interested to find that these 'found-algorithms' included a high proportion of machine learning models geared towards providing a nostalgic kick – generating 8-bit audio, new Pokemon cards, or beating the old Mario games. There's something strange about using a technology that potentially has negative consequences for the sustainability of our future to re-animate something of a nostalgic past[7].

[7] Though machine learning is also used as a tool to mitigate the effects of climate change. For example, see the Alan Turing Institute's 'Environment and Sustainability' programme.

The work’s title emerged from this confluence of rose-tinted glasses and sustainability. The piece is split into two parts, *Rose* and *Green*. During *Rose*, AI is shown learning how to play various video games alongside music generated by AI for video game hardware. *Green* juxtaposes AI-generated Climate Change Pokémon, and audio-descriptions of them, with scientific research on the carbon cost of machine learning. It is set to a background music performed by an AI-generated synth patch.

2.5.2 *Rose*

Rose utilises ‘LakhNES’ (Donahue et al 2019) for its audio. ‘LakhNES’ is a **transformer** designed to generate four-part 8-bit audio, utilising the same specifications as the Nintendo Entertainment System video game console. In an interesting conflation of symbolic and audio AI, it first generates MIDI which is immediately encoded into sound by the four instruments in the ‘NES Ensemble’ (the monophonic instrument voices on the NES soundchip[8]). It is provided pre-trained, so I was only required to press ‘generate’ to receive audio.

From the generations, I isolated a few different styles. There were generations that mostly produced content for the noise generator instrument of the NES Ensemble. I appended these into a base layer of rhythmic noise groove. I divided the remainder of the generations into three groups, each of which corresponded to video footage of AI networks learning to play various games (Mario, Space Invaders, and Hide and Seek). I wrote a systematic pattern that dictated the frequency of switching between these three videos, and accordingly between the three styles of ‘LakhNES’ generation. Towards the end of the piece, the videos are overlayed, as are the audio generations, creating a chaotic aural environment. This was an experiment in collaging audio-generated AI. I noticed that it wasn’t always clear when a ‘LakhNES’ generation was

[8] Comprising two pulse waveform generators, one triangle waveform generator, and one noise generator

starting or stopping, especially towards the end when the frequency of change becomes greater. To make this clearer, and to heighten the theme of nostalgia, I added the sound of a Nintendo GameBoy power-on screen to the start of each audio generation.

While looking for visual found-algorithms I found several that involved teaching an AI how to play video games. *Rose* shows the steady progression of AI learning to become better at the featured video games using these algorithms. Using and displaying a pre-trained model would not have achieved this, since it would only showcase an AI at one point in the training progress. However, training them myself and recording their progress intermittently would have transgressed my zero-carbon no-training tenet for *Rose Green*. I opted instead to use excerpts from extant videos showcasing their entire training process.

The games shown during *Rose* are *Super Mario Bros.*, *Mario Kart*, *Space Invaders*, and Hide and Seek. The first two are learned by 'MarI/O' (SethBling 2015), the third by an unspecified **reinforcement learning** algorithm and the last by OpenAI's 'Gym' (Brockman et al 2016).

These algorithms are not **deep learning** algorithms, which sets them apart from every other AI algorithm in this thesis. They are **reinforcement learning** algorithms, which learn through a process of trial-and-error (Li 2018). An AI is given a task, such as completing a level in *Super Mario Bros.* or hiding from the red team in Hide and Seek[9]. The algorithm will then proceed to retry the exercise continuously until it achieves the goal. This is shown in *Rose* through, for example, 'MarI/O' continually dying in *Super Mario Bros.,* only to get slightly further through the level next time.

[9] Task-setting is not a trivial step and requires innovative thought from programmers to match the goal of the actual game as we understand it with a more statistical goal that an AI can work towards.

Reinforcement learning AI needs no dataset to learn from, only a goal and a means of interacting with the system or game[10].

The learning process was then shown on screen from start to finish, requiring the Hide and Seek game footage to be sped up as the source was much longer than the others. The movement therefore had a natural start and end point.

2.5.3 *Green*

I used two **audio-generative** AI algorithms for *Green*. The first was called 'Synth1GAN' (https://www.thispatchdoesnotexist.com/). 'Synth1GAN' is an AI that creates a new set of synthesizer voices for the Dachi Lab Synth1 instrument, or more specifically a digital VST instrument that emulates the Synth1. It is modelled using a **GAN** architecture, a popular type of AI architecture for visual art creation. I selected my favourite AI-generated synth preset and used it to perform the background music of *Green*. The background music in question is the *Battle Music* from the GameBoy Colour game *Pokemon Yellow*, beginning very slowly and accelerating to normal speed by the end of the movement. Again, this was chosen for its warped nostalgia aspect.

The second was a text-to-speech generator, trained on dozens of famous voices from television, music, politics, and video games. This was called 'vo.codes' at the time, but is now known as 'FakeYou' (https://fakeyou.com/). 'FakeYou' uses **vocal synthesis** to render text as though it has been spoken by one of these voices.

I used 'FakeYou' to read out AI-generated descriptions of Climate Change Pokémon in the voice of David Attenborough, alongside AI-generated images of these invented Pokémon. I thought Attenborough was a good fit

[10] By contrast, solving this problem through deep learning might, for example, involve using an AI to learn from the record of a thousand players who have successfully completed *Super Mario Bros*, in order to infer a solution to the problem. This is, in my opinion, less interesting to watch in real-time.

due to his status as a broadcaster and environmental activist. There was a certain comedy in the juxtaposition of the serious Attenborough with the nostalgic and bizarre AI-generated Pokémon, especially when the AI-generated Attenborough fundamentally fails to pronounce a word. Interspersed with the fake Pokémon descriptions are genuine quotes showing the climate damage of certain AI technologies.

These fake Pokémon images were created through an **image-generative AI**. Specifically, I used an AI called 'BigSleep' (Lucidrains 2021), which allows users to type in a description which 'BigSleep' will attempt to visualise[11].

The text that 'FakeYou' was reciting was generated using 'GPT-2', prompting it with sentences such as 'This is a description of a Climate Change Pokémon' and using its responses. The images were strange enough to maintain interest, while also fulfilling my goal of a confluence of nostalgia, AI, and climate change.

[11] 'BigSleep' is itself a combination of two AI algorithms: 'CLIP', released by OpenAI (Radford et al 2021), and 'BigGAN' (Brock 2019). 'BigGAN' is an image generator and 'CLIP' is an AI that matches images with descriptions. A user can type instructions into 'BigGAN' which will generate images. 'CLIP' will check these images to see if they fit the caption according to its own previous associations of images with captions, and if 'CLIP' does not find a correlation, 'BigGAN' will try again with improved parameters. This, essentially, allows users to type whatever they want to see, and 'BigSleep' will produce something it deems fits the description within a few minutes. Since the model is fully trained already, only requiring a user to type in a description to access the AI's latent space, it fitted with my idea of not training any new models for *Rose Green*. In 2022, this approach became very popular with mainstream audiences online through the models 'Dall-E' and 'Dall-E Mini'.

2.6 Sound of Contagion Project

During the pandemic, I was invited to co-found a research network called 'Sound of Contagion', which used AI algorithms to explore texts throughout history written about pandemics or plagues (www.soundofcontagion.com).

For a 'Sound of Contagion' event at Oxford University in November 2021, I wrote a piece for tenor and synthesizer called *Disc Fragments,* which used both **text-generative** and **audio-generative** AI.

This piece is in seven movements, the first three and last three being somewhat symmetrical – that is, Movement 1 is similar to 7, 2 to 6, and 3 to 5. This arch form was a parallel to the nested narrative form that Chelsea Haith, one of the 'Sound of Contagion' co-founders, assembled from 'GPT-2' texts we had generated for an earlier (2020) stage of the project. The texts for six movements (all except Movement 4) were generated by **fine-tuned** 'GPT-2', fine-tuned upon a dataset of Manchester-based poetry I put together during lockdown.

These texts dictated the form and material of the music. I will discuss the first movement, *At Delphi,* as an example. There was a certain mysticism that intrigued me in this text. The use of 'Delphi', 'the Impossible Him', and 'put the words in your mind' implied to me that 'GPT-2' could be read as imitating an oracle. I wanted the music to create a sense of ritual and of unseen pattern, similar to (for example) Messiaen's *Quartet for the End of Time*. The voice part is constructed from a series of interlocking patterns. Ten pitches repeat, superimposed on a rhythmic pattern repeating every nine notes. The text is simply applied to this pattern. The synthesizer is set to an organ voice and uses chorale-style material generated by AI, using an algorithm developed by Omar Peracha (https://omarperacha.github.io/make-js-fake/). This repeats itself every twenty-three beats (Figure 19).

Figure 19: First 7 bars of 'At Delphi' from 'Disc Fragments' showing synthesizer part derived from AI interacting with isorhythmic vocal line

In this AI generated text, among others, I was struck by its sudden ending. When 'GPT-2' is generating, the user instructs how many characters to generate. Due to its **transformer** architecture, 'GPT-2' does not *plan forward.* It only *looks back* to what has already been generated whenever it generates a new **token**. When it reaches its arbitrary character limit set by the user, it simply stops. It cannot *plan* a 100-character length 'story', for example, because it is not able to *plan forward*. In this movement the music simply stops when the text comes to an end, which does not coincide with the end of any pattern described above. The inspiration for such a technique comes also from composers such as Birtwistle (i.e., *Carmen Arcadiae Mechanicae Perpetuum*) and Edmund Finnis (i.e., *The Air, Turning*).

For the fourth movement of *Disc Fragments*, I trained 'SampleRNN' on a dataset of song, specifically *Winterreise* by Schubert, performed by Dietrich Fischer-Deskau. I chose this dataset because it was relatively short: the shorter the dataset, the quicker 'SampleRNN' is to learn. Since this was my first use of 'SampleRNN' in a piece of music, I predicted I would need to train it several times before it began to produce useful material (i.e, not silence or static). I also chose it because it is also music for tenor, like *Disc Fragments*.

'SampleRNN' is used only in the fourth movement – the keystone movement of the arch form, called *Imitation*. This is a very short movement, lasting only around fifty seconds. I generated several 'SampleRNN' samples and created an electronic soundscape from them by overlaying two on top of one another. I then transcribed some of the noises 'SampleRNN' was making in its effort to imitate a classical singer (Figure 20). These were given to the live tenor, resulting in a game of imitation between the AI-generated song-like material, and the human performer.

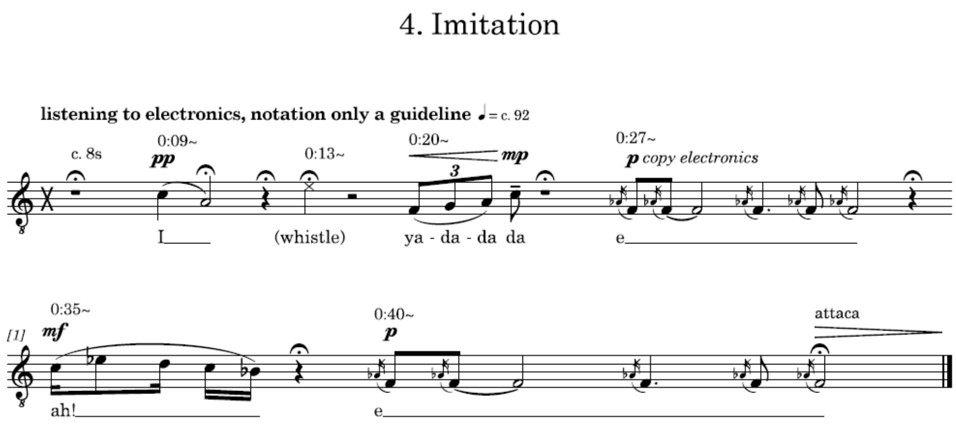

Figure 20: Vocal part for 'Imitation' from 'Disc Fragments'

2.7 Reflection and Future Development

Working on incorporating AI into my creative process through these pieces led me to two kinds of reflections. The first concerns technical progress: how did I find using them, what was or was not useful, and how might the techniques prototyped in this chapter be developed in the future?

The second involves aesthetic concerns: what issues, both musical and extra-musical, are raised through working with these algorithms? How might I address these in future pieces? The most important of these, which I returned to in later pieces, are what I termed:

1. Future and Past
2. Authenticity
3. Structuralism in music
4. Form and musical time as material

2.7.1 Techniques

The three main musical techniques which I codified and included in my practice were interlocking, hidden layers, and collaging. This is in addition to gaining practical skills. Working on these pieces challenged me to create electronics tracks, and my coding abilities were vastly improved by implementing all the algorithms described in this chapter. Both of these new-found skills proved vital for realising my later PhD piece *Silicon* and have been valuable in many other projects since.

I found interlocking an interesting and useful technique for providing new musical material and as a means of reconsidering form. I wanted to use the technique in new contexts, including pieces with larger forces which would allow me more flexibility and creativity when it came to orchestrating the generations. *Three Entistatios* and *Alter* both used 'MuseNet' composing in my style – I also wanted to experiment further with other styles

This is similar to my thoughts on hidden layers. In both ensemble pieces, I enjoyed working with hidden layers and found it a very useful way of challenging my existing notions of what the 'kernel' of a piece should be. My main thought was that I had not pushed far enough into the idea of hidden layers. In future music, I wanted to push myself further and give myself the freedom to be more daring – on reflection, I felt that I had absorbed AI hidden layers *into* my voice, rather than pushed my voice to new places by doubling-down on these ideas. Repeated material was one specific hidden layer I wanted to return to. One example was repeated material: while the AI had a very idiosyncratic and, in my opinion, very interesting approach to repeated material, my own music had tended to obfuscate this.

Collaging was, in my view, not a particularly successful technique when applied to notated music but had more potential as a means of approaching audio-generative AI. Without the performing musician interpreting musical material, the link between AI and audience might become much clearer in this case. Incorporating collaging into an electronics track would also give me a large degree of freedom relating to the spatialisation of AI generations, which might allow any given generation to stand out individually, while also comprising a complex soundscape.

2.7.2 Future and Past

The more I worked with symbolic AI, the more I became fascinated by the relationship between the future and the past, in both AI research and classical music. As exemplified by the *Turing Test // Prelude*, but also the other pieces, a common yardstick for measuring the success of an algorithm is how far it can imitate an existing composer – rather than, for example, whether it can produce entirely new and surprising music. In classical music, I also saw a certain relationship between the future and the past, particularly in orchestral music. New music is almost always programmed next to music that is centuries old, and most of the time it uses the forces and technologies of 19th century Europe. Both AI and classical music, in certain specific ways, use the past to create the future – I was keen to explore this further. I also began to think about the relationship between using past music to imagine the future, and the strengths and limitations of utilising an existing musical grammar to make a wider aesthetic point.

2.7.3 Authenticity

I became increasingly interested in authenticity through my work on the *Turing Test // Prelude*, and the audio-generative AI 'FakeYou', in addition to through two auxiliary projects described in Appendices 1 and 2. I wanted to explore what authentic music is, whether AI can write authentic music (and what type of music that might be), and conversely what 'fake' or 'inauthentic' music might sound like. This was also inspired by comments from performers on *Turing Test // Prelude*, who pointed out the unintuitive nature of AI-generated material when actually putting it under the fingers. Using 'FakeYou' and 'NSynth' (see Appendix 2) introduced me to the sub-field of AI called **style transfer**. I was taken by the creative possibilities of style transfer algorithms and their place in a wider aesthetic commenting on the reality or un-reality of AI-generated music and I returned to them in *Silicon* (Chapter 4).

2.7.4 Musical Structuralism

Using different AI algorithms, particularly symbolic ones, encouraged me to consider deeply what it is that machine learning was *learning* when a computer scientist builds a model to generate music. I was particularly interested in this question in relation to a practice known as dataset augmentation.

Dataset augmentation is a useful approach to curating a dataset for machine learning that is widely used across all fields, not only music. Many algorithms learn best from a large dataset (Radford et al 2019), but sometimes there is only a small dataset available. Dataset augmentation simply applies transformations to the existing dataset to create a larger dataset from it.

In musical machine learning, this is often done through a combination of transposing the music in a dataset into every available key, re-orchestrating the music to more or fewer instruments, or proportionally altering rhythms, thus giving more music from which to learn (Huang et al 2018; Liu et al 2020; McLeavey 2018).

Dataset augmentation in this way overwhelmingly prioritises the relationship between pitches and rhythms, to the exclusion of other musical parameters. Furthermore, even though it concentrates on pitches and rhythms, in fact it is concentrating on the *relationship* between these musical parameters. To the algorithm that has a dataset of every piece transposed into every key there is no fundamental difference between two pitches – only how those pitches interact with other pitches or with other musical parameters to produce music. Dataset augmentation therefore implicitly treats music as an *emergent* phenomenon – the music emerges from the relationship between parameters, and so long as that relationship remains proportionally the same, it does not matter how you transform the music. In my practice, I labelled this as musical structuralism, after the philosophical idea of structuralism (Campagna 2018).

Without labouring the point, it is clear there are many ways that this is often not the case. Even just focussing on pitch, taking the case of the violin as an example, not all pitches are equal. Some are open strings; others harmonics of open strings. There is a bite to low register the fourth string that does not exist in the same place on the first string. Transforming such violin music might keep some relationships intact, but it loses a kind of essence to the sound. This is before speaking of music that exists outside of the twelve-tone equal-tempered system and therefore cannot be transposed without altering its essence even further. That said, it is not true that music can never be a structuralist affair. I find it helpful to consider serialism in this way, though serial composers did not *only* treat music as emergent but rather purposefully explored the spectrum of musical structuralism as part of their technique. This relationship between accepting and rejecting musical structuralism informs music described in Chapters 3 and 4.

2.7.5 Musical Time and Form

Finally, there is the area of musical time and form. To me, one of the more inspiring elements of working with musical AI, and algorithms more generally, is the expanded view of musical time that they offer.

Working with iterative AI that repeats a cyclic learning process, that repeats the same steps over and over with each epoch providing incremental change, was the inspiration for the form of the third movement of *Three Entistatios* and, more loosely, *Alter*. In these instances, particularly *Three Entistatios III*, form was intended to be the most important musical material, as opposed to form being a description or analysis of how other musical materials interact. I enjoyed working in this way and felt I could push the idea further; *Three Entistatios III* is quite short (c. 3 minutes) and the cyclic form is realised through a simple twenty-note cell. I wanted to make more of cyclic and fractal form and returned to the idea in *Chromodynamics* (Chapter 3).

Working with AI also provided me with a new perspective on musical time. AI algorithms do not usually understand time as we hear it, and need to be instructed to generate music in a way that makes sense to the user. In many cases, AI algorithms generate a multi-dimensional latent space that does not immediately exist in time (Caillon & Esling 2021). In others, music is treated more like an image – static in time – and it is only later that the

AI is instructed to ‘read’ this image left-to-right (so to speak), rather than all at once or in some other orientation. Finally AI algorithms, including ‘MuseNet’, may generate music left-to-right by default, but at each moment there is a vast probability tree it chooses the next note from. Some AI can hold many diverging branches of this statistical tree in their memory at the same time, theoretically allowing the user to explore a kind of many-worlds interpretation of the given music, seeing several AI-generated futures simultaneously. As a composer already interested in musical time, working with these algorithms was fascinating. I decided to delve further into these areas in the first movement of *Silicon* (Chapter 4).

3: Developing My Compositional Voice

During 2021 I finished three pieces that built upon and consolidated the musical and aesthetic ideas introduced in Chapter 2. These were *Chromodynamics* for chamber ensemble, *Gravity* for string quartet, and *Warp* for piano and orchestra

Through these pieces I explored much more deeply music relating to algorithmic time and musical structuralism, and they were also influenced by my thoughts on hidden layers inspired by AI-generated music, such as repeated material and silence.

3.1 Chromodynamics

Chromodynamics (2020-21), a 10-minute piece for 11 players, is a musical translation of a physical force - in this instance the strong nuclear force which is mediated through quantum chromodynamics (QCD). While learning about quantum chromodynamics experiments (such as particle collisions at CERN), I became interested in the passage of time for different constituent parts (gluon, electron, proton, etc). They all experienced the same experiment but with time passing at different speeds due to their relative masses (Einstein 1920). I wanted to make a piece of music where the same event (the music) occurred multiple times but in different reference frames of time.

I had prototyped this idea in the third movement of *Three Entistatios* (Chapter 2.2). I wanted to use *Chromodynamics* to further my technical ability to write something using this kind of algorithmic time. Since this piece was much longer than that movement, I felt I had space for developing this idea.

Each reference frame in *Chromodynamics* is shorter than the last and each consisted of the same music. There are 12 reference frames, but only the first 7 are marked in the score as the final 5 are too fast to do so:

1. Neutrino (5 minutes)
2. Electron (2 ½ minutes)
3. Quark (1 minute)
4. Proton (35s)
5. Atom (15s)
6. Collider (8s)
7. Computer (4s)

Each reference frame is the same in that it consists of the same internal sections. There are four sections within a reference frame which I called Collision, Merge, Stretch, and Jets. They can be seen clearly in the first reference frame, beginning around Bars 1, 56, 100, and 118 respectively.

One issue with my use of cyclic algorithmic time as form in *Three Entistatios III* was that there was not enough different, memorable musical material. This made the whole movement feel like a continuous flow of music, rather than the same idea repeating and accelerating. Therefore, I composed these four sections to sound quite different to one another. Collision had fast material generated through mathematical process, shot through with a harsh melody on the oboe (Figure 21).

Merge's principle material was a glockenspiel and crotale ascending sequence (which takes its notes from other instruments' Collision material), accompanied by harmonics and held notes (Figure 22).

Figure 21: Example of 'Collision' section in 'Chromodynamics'

Figure 22: Principal material for 'Merge' sections in 'Chromodynamics'

Stretch was dominated by a sliding melody and alternating chords (Figure 23), and Jet was formed of three iterations of a transposing harmonic sequence and quick flurries of ascending and descending perfect fifths (Figure 24).

Figure 23: Example of 'Stretch' section in 'Chromodynamics'

These aspects remain the same in each reference frame, no matter the speed, making the beginning and end of each reference frame much clearer and focussing my compositional process while writing the piece. I also used a spring coil to mark the beginning and end of each reference frame, an instrument that is heard nowhere else in the piece. This was inspired by Andrew Norman's use of a quarter-tone piano to mark sectional repeats in *Sustain* (2018).

Figure 24: Example of 'Jets' section in 'Chromodynamics'

The reference frames were also the same in that I did not transpose the pitches between them. I wanted the form of the music to be clear and I felt that the constant transposition of the cyclic cell in *Three Entistatios III* contributed to the music sometimes sounding as though it were iterating on an idea, not repeating it.

The reference frames were, however, differentiated through orchestration. I divided the non-percussive ensemble into three groups (Figure 25). In each reference frame, musical material within a group is swapped (Figure 26). Changing the material between the groups allowed me to vary the music enough from start to finish, while still maintaining clarity of form.

RED	BLUE	GREEN
BASSOON / CONTRABASSOON	FLUTE / PICCOLO	OBOE
CELLO	VIOLIN 1	VIOLA
BASS	VIOLIN 2	CLARINET / BASS CLARINET

Figure 25: Instruments in 'Chromodynamics' divided into three groups. Groups are named after colours in reference to the three colour groups of quantum chromodynamics which determine the flavour of quarks and gluons

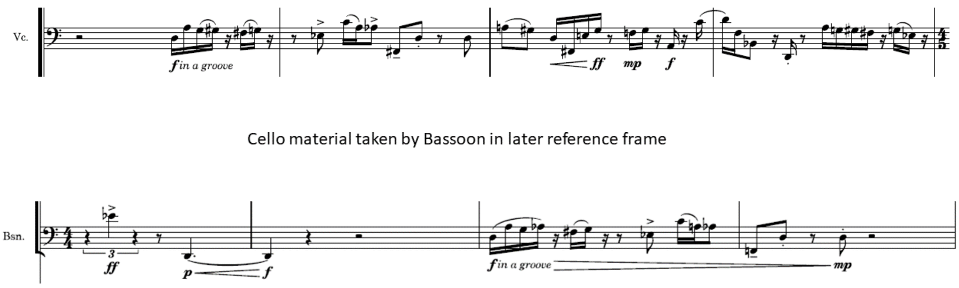

Figure 26: Example of material shared between colour groups in 'Chromodynamics'. Cello material in first reference frame is played by bassoon in second reference frame

Overall, I was happy with my development of cyclic accelerating form in *Chromodynamics*. In future, I would like to try embedding such a cyclic algorithmic time within a larger structure, therefore using it as a means of getting somewhere new, rather than the central idea to the entire work.

3.2 Gravity

3.2.1 Overview

Gravity (2021) is another piece translating fundamental forces of nature into music, alongside *Chromodynamics*. One way of looking at the piece is that the five movements map the chronology of scientific explanations for the force of gravity. The movements are as follows:

1. Earth & Sky
2. Spheres
3. Universal Law
4. Curve
5. Graviton

Gravity explores musical structuralism in two ways: through tuning and through what I termed ‘arrays’. It also explores repeated material, building upon the hidden layers explored in Chapter 2.

3.2.2 Tuning and Musical Structuralism

Of particular interest to me while writing this piece was the research undertaken by English music theorist Thomas Salmon at the turn of the 18th century. Salmon demonstrated an alternative tuning system at the Royal Society in 1705 based upon Newton’s work on the theories of Universal Gravitation and the nature of light while isolating from the Great Plague (Wardhaugh 2017).

This was attractive for three major reasons. First, it seemed fitting to utilise a tuning system developed in response to Newton’s work on gravity for this work translating gravitational theories to sound. Second, there is an obvious parallel between Newton’s work in isolation, and my own work in isolation on this music during the 2021 Covid-19 lockdown. Third, most importantly, there seemed a rich parallel to be drawn between this tuning system and my ideas on musical structuralism, described in Chapter 2.

Like any justly-intoned tuning system, Salmon's proposed temperament is not infinitely transposable because it encounters the issue of the Pythagorean Comma (see Hubbard 2022 for explanation of the Comma). Similarly, any given note does not have a fixed frequency to which it always correlates, but rather a dynamic frequency based on the key the music is in. This is quite unlike equal temperament, which treats notes as the same (in terms of frequency and intervallic relationship to other notes) in all contexts. This property of equal temperament allows the dataset augmentation practices described in Chapter 2.

Music written using systems like Salmon's cannot be transfigured or manipulated without changing the essence of the notes. This made it useful for me in my attempt to critically consider musical structuralism's role in my compositions.

I then had to decide how to implement this idea in practice, in a way that might be effective when heard. In doing so I referred to the music of composers well practiced in employing microtonality for new music, such as Lawrence Dunn (*Set of Four* 2017) and Robin Haigh (*Grin* 2019). There is also the issue of tonality. Salmon's system is designed for tonal music, but my music is not usually always traditionally tonal.

To solve these problems, I created a fixed chromatic scale, based on Salmon's work (Figure 27). When directed to play using 'Salmon-inspired' tuning[12], players would not need to re-tune every note depending on the tonal context, but instead would use this fixed chromatic scale (from here: Salmon-scale) version of that note. This fixed chromatic scale was based on D major which aligned with the open strings of the quartet.

[12] I call this Salmon-inspired because it is not exactly what Salmon wrote in his paper 'A proposal to perform musick in perfect and mathematical proportions' (Wardhaugh 2017). In practice it would be very difficult to precisely perform Salmon's system on modern instruments (his own demonstration in 1705 had specifically re-tuned viols). The fixed Salmon-scale therefore was based on an approximation of Salmon's tuning system, allowing the quartet some margin for preference during performance

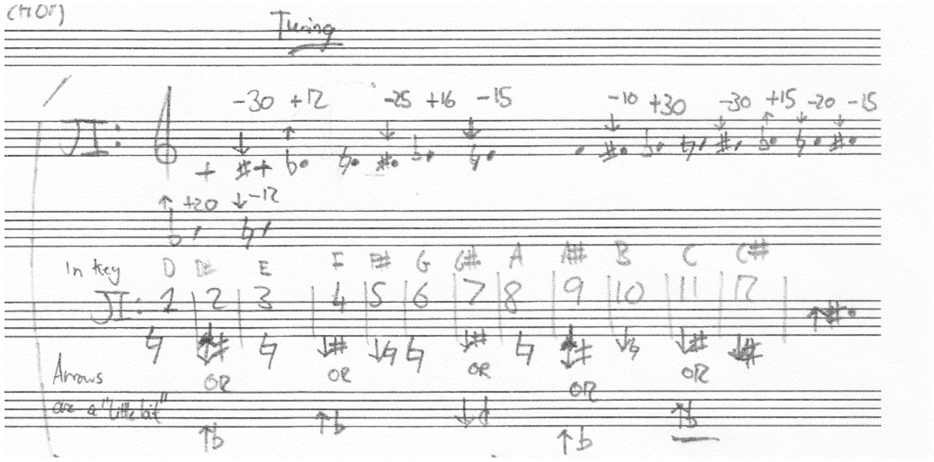

Figure 27: 'Salmon-Scale' from composition notes to 'Gravity'. Top line shows deviation by cents from equal-tempered D major scale. Bottom line shows approximate notation used in quartet, where an arrow represents a slight shift in tuning and quarter-sharp/quarter-flat symbol represents major deviation

During the piece, the Salmon-scale is employed in several ways to transform infinitely transposable material into material with a kind of internal essence. Being fixed in D major lends the scale a certain magnetism towards that tonality, even when (or especially when) the music is not tonally centred on D or is fully atonal.

For example, the scale is first used towards the end of *Earth & Sky* (Figure 28). This section is a simple cycle of fifths, moving from F# major through to D. Using the Salmon-scale, the cycle of fifths was pulled towards D major through the warping of intervals in other keys, even though in equal temperament a cycle of fifths is infinitely transposable, and therefore (in my view) tonally neutral without further context.

Figure 28: Salmon-Scale imposed on a cycle of fifths at the end of 'Earth & Sky' in 'Gravity'. Salmon-Scale warps the 'neutral' cycle of fifths towards D major. For example, the G#-D# open fifth in Bar 102 is so flat that it is better to notate the D# as a quartertone, while the F-D open fifth in Bar 105 is slightly sharp.

The tuning can also work in competition against the harmonic material of the music. From Bar 34 of *Curve*, for example, a sustained tone E is established across the quartet (Figure 29). At this point, the Salmon-scale is dictating the tuning of each note, though the material has no tonal centre. In regular tuning, therefore, the sustained E might have become (aurally) a tonal centre to focus upon, in the absence of any other tonal argument. The Salmon-scale's magnetism towards D major, however, forces me at least to contextualise the E as an anticipatory note. The music then moves to a C major 10th in Bar 40, skipping over D. The root of the Salmon-scale is not heard, forcing the music (in a harmonic sense) to continue until it is (in Movement 5). Where the Salmon-scale gave the cycle of fifths in *Earth & Sky* a sense of a place to rest amongst infinite transposability, the scale here forces the music to *not* rest.

3.2.3 Arrays and Musical Structuralism

Figure 29: Salmon-Scale working against harmonic content of music in 'Curve' from 'Gravity'

The use of tuning throughout *Gravity* is an example of what I called an ‘array’ while composing the piece. An array, in this context, is a series of positions within one musical parameter. For example, the musical parameter ‘Tuning’ has been the focus of the discussion above. One position would be the Salmon-scale, another equal temperament, a third might be natural harmonics (the focus of the second movement *Spheres*). I wanted to reduce the essence of my music to as few arrays as possible, and then explore the combinations of different positions on these arrays through the piece. For *Gravity*, my arrays were: tuning, timbre, texture, and time (Figure 30). This is quite a structuralist (according to my definition) approach to music-making: the music itself was often an emergent result of the interaction between arrays.

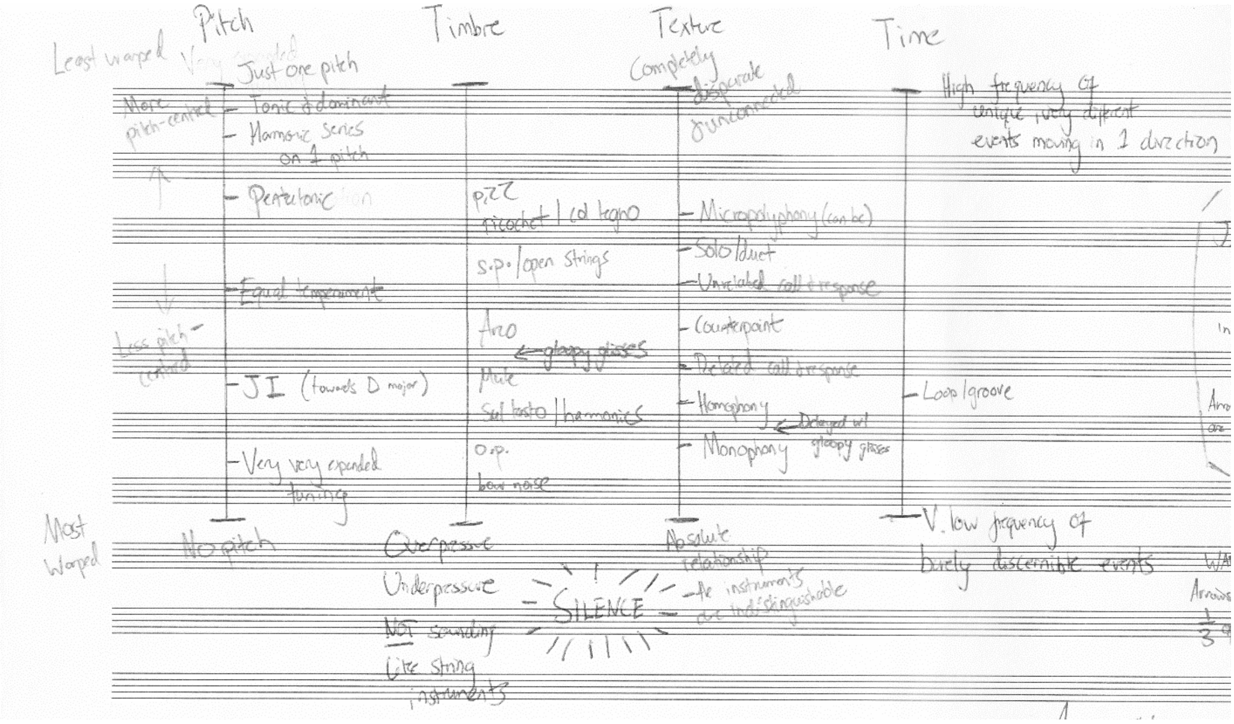

Figure 30: Arrays for tuning (pitch), timbre, texture and time in 'Gravity' planning. For example, the maximum texture was defined as where each member of the quartet is completely separate from the other three, with the minimum being each instrument is playing exactly the same music. Similarly, a minimum timbre was defined as each instrument sounding exactly the same. I found that the best way to achieve minima on all the arrays was total silence

I was interested in translating Einstein's theory of General Relativity in some way into music, especially during the fourth movement *Curve*. In general relativity, time is not a blank canvas upon which the universe is painted but is rather a malleable dimension alongside the three spatial dimensions that can be warped (Einstein 1920). This was of particular interest to me, due to my renewed interest from working with machine learning algorithms in exactly what might constitute musical time. General relativity seemed to be a fascinating way of viewing the universe through a structuralist lens (while also being an integral part of the story of gravity).

One of the main problems with general relativity that leads physicists to believe it is incomplete is its inability to sensibly deal with what happens at extreme scales, particularly extremely small scales, close to infinity or zero (du Sautoy 2016). Once I had made my arrays, I wanted to push them to the same extremes that theoretical physicists pushed general relativity and observe the resulting music. To do this, I tried to theorise what a 'maximum' and a 'zero' was for each array.

To do this with the time array, I needed to create a working definition of musical time. This is a contentious area and I do not claim my definition holds for other peoples' music, or even for my own other pieces. The definition I settled upon was:

The ***frequency*** *of* ***discernible*** *events that* ***traverses*** *an* ***array*** *in one direction*

In the interests of brevity, I do not discuss this definition and its ramifications here but provide that discussion instead in Appendix 4.

I wanted a moment in the piece where all four arrays were as close to 'zero' as possible. This can be heard in *Curve* from Bar 189 to the end of the movement. The texture moves from four part harmony to monophony between Bar 195 and 201. The timbre of each instrument coalesces onto rough scratch tones by Bar 217. The 'tuning' of the instruments is gradually quantised between Bars 189 and 215, moving from 12-tone Salmon-scale to just one pitch being available. These are all the extreme ends of the tuning, timbre and texture arrays. Once they have reached these extremes, they do not change again for the rest of the movement. Thus, according to my definition of musical time, the music should feel close to timeless from Bar 217.

I found it did feel close to timeless, but not entirely in stasis. I postulated this was because the musical events were not quite indiscernible from one another, as they should be according to my definition of musical time.

String instruments do not all sound alike, and a G quarter-sharp scratch tone on a cello does not sound *exactly* the same as a violin, or one violin the same as another. They have a slightly different timbre, and therefore the timbre array is not at its minimum (which is defined as *exactly* the same). Because they have a slightly different timbre, the instruments are therefore also not in perfect textural unison: they are (audibly) not playing *exactly* the same thing. The only way to ensure each instrument sounds *exactly* the same, and therefore truly achieve the 'zero' of the four arrays, was to have each 'play' total silence. This is why the movement ends in silence, and it is also why I placed a dynamic underneath the rest. The rest is not a 'lack' of music, but it is rather the four arrays pushed to their absolute extremes. It is a logical conclusion of my personal translation of general relativity into music, a kind of sonic event horizon being crossed.

3.2.4 Repeats

In *Gravity* I returned to the idea of finding new ways to consider the role of repeated material (Chapter 2). I wanted to explore further the way AI algorithms will become stuck in a certain place, repeating the same music (or text) for a very long time before breaking the pattern after some seemingly-arbitrary number of repeats (3, 50, 1000) and moving on.

An example is in *Universal Law,* a movement which is a slow coming-together of the four instruments. They each play their own melodic line in different rhythms, each time transposed away from the last. Gradually these pitches and rhythms come together by Bar 51 into a series of chords (Figure 31). Here I wanted to capture a sense of imperfection, a system that seems elegantly simple but ultimately is unsatisfactory (like Newton's Law of Universal Gravitation). Channelling machine learning's approach to repeated material seemed a good way to portray this uncanny imperfection. In a landscape that has been constantly transposing, it is intended to be strange to hear the same music multiple times. Similarly, the dynamic level jumps immediately from *forte* to *pianissimo* while the rest of the movement has used exclusively gradual dynamic shifts. Metric dissonance is introduced through the use of a 2/6 bar (the length of two crotchet triplets), which seems to place the repeat halfway through a triplet.

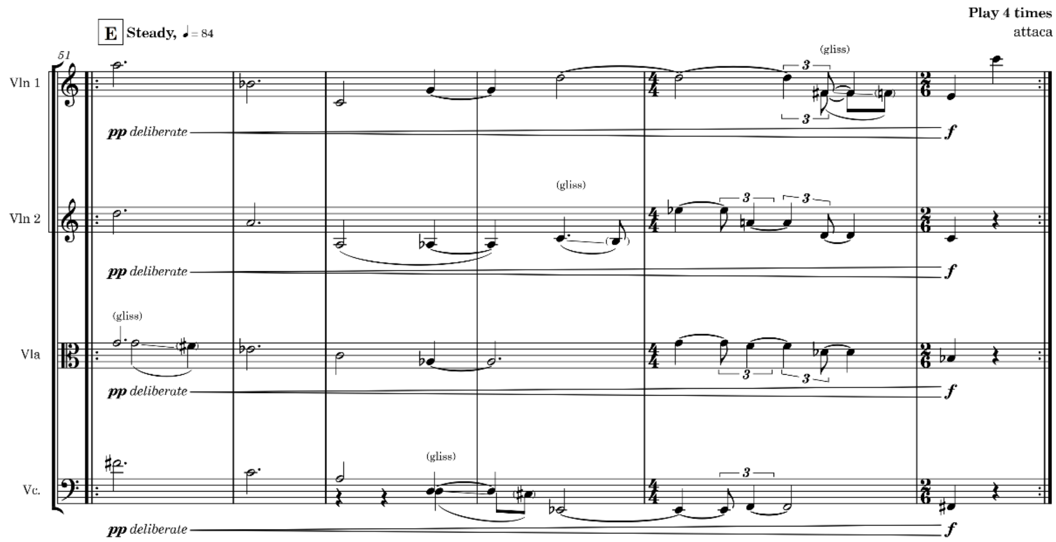

Figure 31: Use of repeated figure in the final six bars of 'Universal Law' in 'Gravity'

3.3 Warp

3.3.1 Overview

Warp (2021) is a 12-minute piece for solo pianist and symphony orchestra. Programmatically, it attempts to translate a particular solution to Einstein's field equations into sound. This solution, first proposed by Miguel Alcubierre, suggests it is possible to exceed light-speed travel by warping the fabric of spacetime around a spaceship, while the spaceship itself is travelling at normal (i.e., possible) speeds (Alcubierre 1994). In *Warp* I imagined the solo piano to be the spaceship, and the orchestra a kind of fabric that evolves in complex ways around it.

Warp is a partial continuation of my work on *Gravity*. Warp applies the notion of the 'array' to more instruments within the orchestra, expands the idea of 'musical fabric' translating the spacetime fabric of general relativity into music, and develops my desire to closely consider musical structuralism.

3.3.2 Arrays

I wanted to push my work on creating and utilising musical 'arrays' further than I had in *Gravity*. Throughout *Warp*, I wanted elements of the music to feel stretched or compressed, particularly in relation to the speed of material and its timbre. While planning the piece, I designed a timbre array for each instrument type in the orchestra, the two extremes of which being sounds I considered to be stretched and compressed (Figure 32). Music that fell exactly in the centre of the timbre array might be considered 'neutral' in the context of orchestral repertoire. I divided the orchestra in half spatially and assigned one half to be 'stretched' and one half to be 'crushed' throughout the piece.

Example Timbre Array - Strings

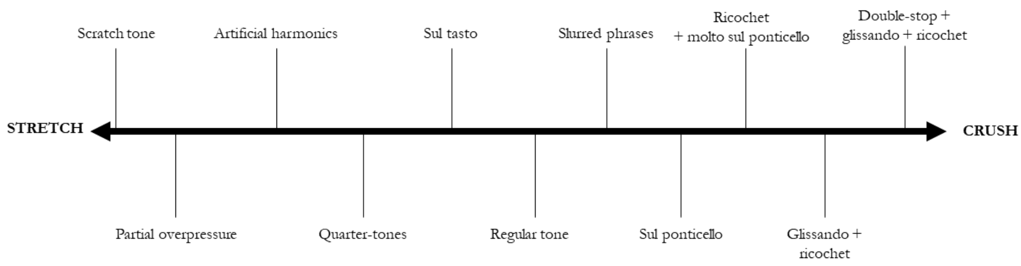

Figure 32: Example of a timbre array used during 'Warp', for string instruments

'Stretched' sounds were generally those that were intrinsically *unstable* (outside of the full control of the player) and those that audibly distort a pitch to the point of it being unrecognisable. I imagined zooming into an image to the extent that I could only see individual pixels, not the image as a whole, or to the level of quantum mechanics where reality seems to be based on probability. Examples include scratch tones and quarter-tones for the lower string instruments and the use of the superball mallet on the timpani.

On the other hand, 'crushed' sounds were those that tended to embrace a wide spectrum of sounds within one action. I imagined zooming out from an image until the distinction between two neighbouring colours are blurred, or pressing all the colours of a rainbow together until it all appears white. This included, for example, the use of sul pont for the upper string instruments to sonically unlock more of the harmonic series and multiphonics in the woodwind.

These timbral arrays were directly tied to a 'speed' array. The further towards being 'crushed' an instrument was, the faster it played, and the inverse for 'stretched'. While this was usually achieved by relatively simple manipulation of rhythm, tempo and metric modulations, there were some places where the timbre and speed arrays interacted (Figure 33). Considering ‘speed’ instead of ‘musical time’ (as I did in *Gravity*) appeared to, at least in my work, reveal a key difference between speed and time. The speed array in *Warp,* like the tuning, timbre, and texture arrays in *Gravity,* became intermingled with other arrays when pushed to its extreme. This was not the case for musical time, which had a different relationship with the other arrays (see Appendix 4).

Figure 33: Marimba part at the point at which getting faster (on the speed array) necessitates a distinct position on the timbre array (glissandi) as the only way to play the musical material at the correct speed. Wide glissandi have their own unique timbre, showing the confluence of the two arrays at their extreme ends

3.3.3 Fabric and Musical Structuralism

While the arrays were to provide the basis of the orchestra’s sonic journey throughout the piece, they are fundamentally modifiers to musical material, not material itself. A player can’t simply play fast or slow – they must play *something* quickly or slowly. I therefore composed a kind of musical fabric through which *Warp* could be woven. This fabric needed to be simple enough that the journey along the timbre and speed arrays would be significant, but not so simple that the 12-minute piece would become overly repetitive. Furthermore, I wanted the fabric to be quite different to the initial material of the solo piano, so that the two would stand in a kind of direct opposition that could be resolved later in the piece.

I decided to base the musical fabric of *Warp* on ascending scales. When designing the scale(s) that would form this fabric, I mainly considered three aspects of scales:

1. The root of the scale
2. The number of notes in the scale
3. The intervals between these neighbouring notes

After two prototype fabrics that weren't satisfactory, I settled upon the following patterns to determine these three aspects.

1. The root of the scale was given by a 7-note wedge-shape whole tone contrary motion scale (Figure 34) that transposed itself up a 5th every 7th scale. One side of the orchestra (crushed) played the top notes in the wedge while the other half (stretched) played the bottom notes, though this only becomes obvious after the two sides diverge from one another temporally.
2. The number of notes in the scale was determined according to an 11-part metric pattern (Figure 35).
3. The intervals between the notes were determined by processing through 8 separate scales, each of which was distinct from the others through modifying the 3rd, 4th or 7th notes in the scale (Figure 36).

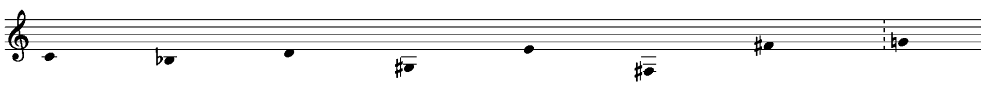

Figure 35: 7-note wedge determining the root of scalic material in 'Warp'. Repeats at the fifth

6 – 5 – 7 – 4 – 8 – 3 – 3 – 8 – 4 – 7 – 5

Figure 34: System determining number of notes in each scale in 'Warp'

1. Major 3, ♯4, ♭7
2. Major 3, ♮4, ♯7
3. Minor 3, ♯4, ♭7
4. Major 3, ♯4, ♯7
5. Minor 3, ♮4, ♯7
6. Minor 3, ♯4, ♯7
7. Major 3, ♮4, ♭7
8. Minor 3, ♮4, ♭7

Figure 36: System determining the quality of intervals within the scales in 'Warp'. Repeated itself indefinitely

Since it moves through so many harmonic areas so quickly, the fabric has no real tonal centre at any point, though one can be artificially constructed through sustaining notes. This gives the solo piano a role to play in establishing the harmonic framework of the work.

Before the two sides of the orchestra diverge temporally at Bar 78, I also added an additional parameter: the length of each note in the scale. This rhythm oscillated between quaver, dotted quaver, triplet crotchet, and crotchet. In conjunction with the *number* of notes in the scale, this gives rise quite naturally to bars with unusual meters (Figure 37). Having such meters built into the fabric of the piece provided an interesting challenge while composing the piano part, forcing the solo music to warp itself in ways that I would not have done through intuition alone.

Figure 37: Unusual meters emerging from combination of systems within 'Warp' musical fabric. The 1/6 bar emerges from the number of notes in that scale being 7, and their length being crotchet-triplet.

Part of the reason for developing and using the musical fabric described above in conjunction with arrays for the orchestral instruments was to investigate how musical structuralism could be either rejected or purposefully embraced within a piece.

So far, I have shown how I intentionally utilised ideas that, to me, are examples of musical structuralism. The phenomenon of emergent music is one, where the sonic surface of the piece is the inevitable result of the underlying forces governing the piece. Another is infinite transposability: the musical fabric described above is an example. Having no start or end point, musical material within the fabric is only identifiable through its relationship to other parts of the fabric.

I also wanted to push the piece to the other side of the coin, to rejecting structuralism and embracing essence. What exactly constitutes musical essence is a very large question that's well beyond the scope of this analysis. In this instance I wanted to define it as some part of the musical sound or idea that can only exist in that context. If the musical idea were changed, or orchestrated, or transposed, and so on, it would cease to have this part.

I was helped here by the timbre array I generated earlier in the compositional process (see Figure 32). While the array itself was designed to explore structuralism, the very ends of the timbre array were invariably extended techniques (being far from a 'neutral' sound). Extended techniques are, for the most part, specific to the instruments they are played on; the concept of the scratch tone doesn't really make sense on a saxophone, for example. So by the time the music reached the end of the timbral array, around 2/3 into the piece, there is already an 'essential object' built out of here, by my definition. This can be contrasted to the orchestra at the beginning, where the orchestration of the fabric was extremely fluid and did not demand a specific instrument. I pushed this ‘essential object’ a little further by the end of the work by allowing a harmonic area to emerge out of the open strings (Figure 38). Open strings have some element of essence to them due to their integral physical function of the instruments, and allowing a harmonic area to finally emerge is a rejection of the whirlwind harmonic framework earlier in the work. I pick this idea up in the 3rd movement of *Silicon*, applied to the entire orchestra.

Figure 38: Harmonic area emerging from open strings at the end of 'Warp'

For the solo instrument, I attempted the reverse. The music begins in an intuitive compositional style, with the only dictates upon it being the metric interruptions described earlier. As the piece progresses, the same harmonic progression repeats itself in the solo instrument, except that each time this harmonic framework becomes clearer. By the time the orchestra has reached the stage of their ‘essential object’, the piano has been stripped down to its core components of chords and repeated rhythms. During the cadenza, this is stripped down further: the music begins transposing itself both harmonically and metrically (each repetition being one step slower than the last) until it reaches the same harmonic area as the open strings. I wanted the piano to have an inverse journey to the orchestra.

4: Silicon

4.1 Introduction

This chapter is an analysis of my piece for orchestra and artificial intelligence, *Silicon*. It develops several technical and aesthetic interests and concerns that I had developed over working on the pieces already discussed in this commentary. These include symbolic-generative AI, audio-generative AI, style transfer algorithms, the integration of electronics with classical ensembles, AI and authenticity, the relationship between the future and the past in AI research, AI and musical structuralism, form, and musical time.

Throughout my PhD I have transitioned from music *using* AI to music that is also *about* AI, from a wider, social perspective. Overall, I have become more interested in my work having relevance to the world it is in, and less interested in only writing purely abstract music, which was my practice before beginning work with AI.

I've found AI an excellent topic to respond to for this reason. We interact with AI all the time. It serves me advertisements for upcoming concerts it has learned that I like (Mogaji et al 2020). It changes the directions on my satnav based on live information from other drivers' phones (Lau 2020). It summarises an article I can't be bothered to read, corrects the grammar in an email I send before I've had my morning coffee, recommends new films for me to watch (Tatalovic 2018; Laksnoria 2018; Gomez-Uribe & Hunt 2015). In short, it is everywhere, but more importantly it is, or could be, *anywhere* (Cox & Riis 2018). That's the reality of a world that increasingly relies on algorithms to command hidden infrastructures that support society.

In contrast to those examples already mentioned, we have also seen recently that AI can accidentally replicate harmful biases with recent cases including discrimination against women in the workplace (Dastin 2022), failure to recognise ethnic minorities in passport recognition software (Leslie 2020), and the automatic promotion of extremist social media groups to those already at risk of radicalisation (Hao 2021).

As a technology, AI also forms an integral part of wider issues that encompass many other technologies and social questions. The proposed 'metaverse', hypothesised to synthesise physical and digital experience seamlessly, relies upon AI in concord with other advanced technologies such as augmented reality and wearable technology (Lee et al 2021). Many researchers are focussed on AI as an essential tool to mitigate climate change, mobilising it alongside more well-known political and social arguments (Cowls et al 2021). Climate change is the principal example of what Timothy Morton calls a "hyperobject" – an idea so abstract and vast that it is difficult or impossible for any individual to grapple it in its entirety (Morton 2013). Like the case of climate change, it is entirely plausible that AI, which can absorb and analyse truly vast quantities of data, will become a critical tool in understanding other hyperobjects – and may become a hyperobject itself.

Some of these questions can be explored in an orchestral space, and my piece *Silicon* is a first attempt to do that. The relationship is more than one-way: it is not simply using orchestral music as a vessel to comment on a society dependent on AI. AI can, in return, offer the orchestra new perspectives, technologies, and approaches to music-making that did not exist even a decade ago, even at the same time as certain actors within the music industry work towards replacing human musicians with indistinguishable AI performances. AI is both dangerous and full of potential, both good and bad, both hidden behind-the-scenes and on full display as a corporate buzzword. It has these dualities that are both alluring to artists and highly relevant to the orchestra, which is also, I would argue, partially reliant on a foundation of dualities and contradiction. Three particular dualities have stood out to me as important for discourse around designing and understanding AI, which are also areas that I believe musicians and orchestral institutions already give a great deal of thought. I have termed these:

1. Future and Past
2. Fake and Real
3. System and Secret

Broadly speaking, I have tried to explore each of these dualities in their own movement of *Silicon*:

I. *Mind*
II. *Body*
III. *Soul*

Each also develops other ideas already touched upon in this commentary, reached through work on earlier pieces of music. These ideas are both technical (i.e., my practice of utilising AI in the creative process) and aesthetic (i.e., what kind of music do I want to make in response to working with AI?).

4.2 Silicon Mind

4.2.1 Future and Past

The first duality I worked into *Silicon* was the relationship between the future and the past. Specifically, this relationship is one of legitimacy: how does the past legitimise the future in AI research and in some orchestral music, and how can this form the basis of a new piece of music?

As discussed in Chapter 3, I have found that many contemporary AI algorithms use existing music both as a dataset from which the AI learns the rules of music and also as a yardstick against which to judge the quality of generations. Marcus du Sautoy (2020) writes that ‘Bach is the composer most composers begin [learning] with, but he is the composer most computers begin with too’, and indeed Bach is often the choice of dataset and generation for much recent research (e.g., Hadjeres et al 2017; Fang et al. 2020; Whorley & Laney 2021). The more indistinguishable from Bach, the more successful the AI is deemed, as I played with in *Turing Test // Prelude*. This might lead to the impression that one of the main uses of generative AI might be to complete unfinished pieces by dead composers, as indeed we have seen in recent years with AI ‘completions’ of Beethoven and Schubert (see Goodyer 2021).

Relatedly, several recent AI algorithms outside of academia have been developed that automatically generate music for soundtracks or other media uses (Langkjær-Bain 2018; AIVA 2021). Crucially it is designed to replace composers who write this music – not necessarily to create more interesting music. Once again, AI needs to sound like what already exists as much as possible, in this case to cut costs of human labour.

This leaves us in the situation that this technology, which promises the future, is often looking to the past to prove its legitimacy (rather than proving its usefulness through, say, generating entirely new and surprising material). Of course, there are exceptions to this rule as some researchers do place genuine novelty and creativity at the heart of their research. I wanted to write a piece that leaned into this past-focussed view more heavily than I had done previously and thought the orchestra was a good vessel through which to do this.

Classical music, and by extension the orchestra as an institution, is well-known for using the past to create the present and the future. The performance of established music is perhaps the genre's defining trait, which is evident in the programming of symphony orchestras (Donne, Women in Music 2020; Gotham 2014). Why do we do this? My most optimistic view is that it's because we believe that ideas from the past can have something to say in the present – something beyond merely being a benchmark by which to judge technical progress. Conversely, modern composers often use references to older music, or different genres of music, to make exciting and fascinating musical arguments.

Silicon Mind is scored for a double wind orchestra with a few additional instruments. There are no electronics; the relationship between orchestra and AI is realised solely through the interpretation of sheet music by live performers. It has fewer instruments than the other two movements because it was originally to be premiered in a separate concert, which was cancelled due to the pandemic. I kept the instrumentation for the final version because I felt using a Classical orchestra remained appropriate for this music. In this movement, I wanted to explore this idea of future and past and develop an approach that uses existing styles of music to make my own musical arguments. I also wanted to develop my ideas on AI-inspired form and musical structuralism.

To explore these ideas, I returned to 'MuseNet'. I was very familiar with this algorithm at this point – similar, in fact, to my familiarity with my physical instruments. I had a good idea of how to use it and how to avoid it doing things I wasn't currently interested in. For this movement, I instructed 'MuseNet' to generate in the style of Mozart. This suited the overall theme of creating something new out of something old and allowed me to showcase the algorithm when performing at its intended task (stylistic composition) as well as when pushed into new places. Choosing Mozart also fitted with the Classical orchestra format.

4.2.2 Sonata Form

I felt that it was appropriate to house the 'MuseNet' Mozart material within a form the AI would expect, at least at first, so I chose to compose the movement using sonata form as a departure point (Figure 39). I was interested in exploring and showcasing the change from familiar to uncanny, and in slowly polarising the AI-generated material until it ended up in a very unfamiliar sonic landscape. As mentioned in previous chapters, I am fascinated by the approach that AI takes when repeating material, and how different this is to how musicians normally understand repeats. There are two types of structural repeats in a Classical sonata – the exposition repeat and the recapitulation – and I was interested in the relationship between these expected structural repeats and AI-inspired repeats.

Form is also distinct from musical time. In *Silicon Mind*, more than any other piece discussed so far, I was interested in bringing AI-inspired ideas to musical time. Having a set form and applying musical time manipulation to it is less needlessly complex and, I think, more audible than inventing a form and then applying such manipulation. Beginning a multi-movement orchestral work with a sonata form also seemed a good way of making my wider point about the utility of AI in the creative process *beyond* projects such as 'completing Beethoven's 10th symphony'.

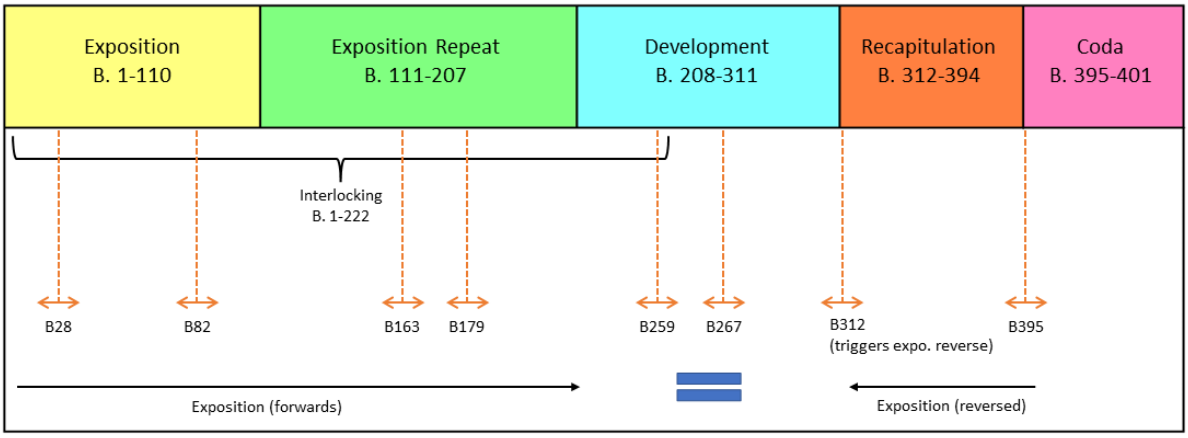

Figure 39: Structural diagram of 'Silicon Mind'. Top shows sonata form relation. Section containing interlocking show with bracket. Red arrows denote locations of axes of reversal. Bottom shows where exposition is heard forwards and, later, reversed.

4.3.2 Interlocking

Continuing from previous pieces, the interlocking in *Silicon Mind* consisted of alternating sections of human-composed and AI-composed music, which can be seen in Figure 39. More specifically, I used AI-generated music in:

- Bars 12-24 (falling arpeggiac development)
- Bars 33-44 (low instrument Bb material with upbeat)
- Bars 100-106 (continuing from 'Silicon Scale' – see later)
- Bars 118-128 (falling arpeggiac and Bb material with upbeat)
- Bars 201-222 (low instrument diads)

Through using 'MuseNet', I developed the interlocking technique I had trialled previously in four ways which I felt marked an improvement: polarisation, development, orchestration, and reversion.

4.2.4 Polarisation

Having attended many conferences and other events in the musical machine learning field, I was aware of how much researchers tend to cherry-pick their results to prove the efficacy of their AI algorithm, discarding a large proportion of generations to show only the best few. As previously discussed, this efficacy is often shown through an algorithm's ability to imitate Bach or a similar composer. This seemed to me to be both pointless, as we already have music composed by existing composers, and (in the worst cases) duplicitous because the usual tendency of an algorithm was passed over in favour of occasional generations that, while happening to sound more stylistically accurate, were not truly representative of the AI in question. I wanted to do the opposite, and cherry-pick results that did *not* sound like Mozart, and instead elected to retain the generations which I found unusual, uncanny, obsessive, and bizarre. Essentially, I wanted to ask 'does AI *really* sound like Mozart?' and 'do you *really* want it sound like Mozart?'.

Selecting which to retain felt a more creative act in this work than previous iterations of interlocking, because I was rejecting the idea of 'best' and instead comparing each generation on their own merits. Once I had selected a generation, I repeated the procedure, sometimes composing a little bit in between (like *Three Entistatios II*). This feedback loop of selecting the most unusual generations, rather than the most stylistically fitting, and then instructing 'MuseNet' to continue those generations very quickly polarised the work. It becomes quite strange, quite quickly, and in this way was reminiscent of the work with text on *Alter* (Chapter 2), which also went down a strange rabbit-hole. This might not be coincidence, because 'MuseNet' has a very similar architecture to 'GPT-2' (used for the text in *Alter*). It is possible they are both susceptible to polarisation through feedback loop in the same way. Some of this strangeness comes from me, of course, because I had always intended the work to go from 'normal' to 'uncanny', but much of it certainly stemmed from AI-generated ideas.

I was fascinated by these generations, because it felt clear to me that 'MuseNet' understood them to be, in some way, as stylistically accurate as the generations that did (to my ears) sound like Mozart. It felt like there was a strange, almost alien, methodology at work under the surface, which was far more interesting than a computer showing it can learn rules we already know concerning harmony, counterpoints, and voice-leading. I wondered if the learning machine had discovered a deeper ‘style’ that underpins this music, which does not necessarily relate to the surface-level features of the music.

I did not only use one ‘MuseNet’ generation responding to a given prompt. As will be discussed later, I sometimes showcased one generation before reverting to the start of the generation and choosing a different one. This allowed me to show two wildly different responses to the same musical idea and gave the music a sense of unpredictability. This marks a different approach to previous interlocking, where I would always choose *only* one generation from a set and continue from there. This also allowed me to orchestrate two responses to the same prompt in different ways (Figure 40).

Figure 40: Two different orchestrations of material separated by ratchet (Bar 35) which reverts the music to the beginning of the phrase

4.2.5 Development of AI-Generated Material

I used 'MuseNet' only in the first part, mostly covering the 'exposition', of the piece. I then developed these ideas in various ways throughout the rest of the work. This is quite different to, for example, *Alter*, where the 'MuseNet' generations continue until the end of the work, allowing no room for me to work on development. While most of these are evident from the piece itself, I will provide some examples:

1. The ’MuseNet’ generation beginning Bar 28 was utilised in several ways across the piece, including being transformed into the material at Bar 50, accelerated and microtonally transposed in Bar 119, and transformed into the rapid descending figuration in Bars 149-160. This figuration was itself developed through reversal, as outlined above.
2. Bars 111-133 (the exposition repeat) develop various ‘MuseNet’ generations through acceleration and microtonal transposition.
3. The ‘MuseNet’ Bb upbeat generation first seen at Bar 33 was transformed in various ways throughout the piece, especially during Bars 290-312 where it is recontextualised as a bass layer (rather than a melody), in counterpoint with an inverted version of itself. Using it in this way allowed it to link directly to the Bb (trombone; cello) in Bar 312, which begins the reversed wide scale reprise.
4. The strings, tuba, and bass trombone material in Bars 207-238 was generated by ‘MuseNet’ and transformed through elongation and transposition (down several octaves). It was then extended into the cadential figuration found in the string from Bar 239, and the whole section made microtonal.
5. The percussion and harp material in Bars 228-254 combines the descending arpeggaic idea from Bar 50 with wide scale notation. Here, I took notes from the wide scale corresponding to the intervals of the arpeggios in Bar 50 and created chords around them through applying set intervals to those notes in rotation. The top notes of these percussion chords were transposed up one or two octaves to become the solo violin and wind material from Bar 267-312.

4.2.6 Orchestration

One recurring question when using ‘MuseNet’ is how to deal with the issue of orchestration. As described in Chapter 2, ‘MuseNet’ has limitations on how many instruments it can generate for, and its ability to understand what those instruments are. Like *Three Entistatios*, I chose to provide short score reductions of orchestral music to 'MuseNet', receiving the same back. I would then orchestrate these later. Unlike *Three Entistatios*, I had an entire orchestra of options to choose from, and so was able to devote more attention to the intricacies of orchestration.

Usually, my approach to musical material has been intertwined with orchestration; the idea exists *as* its orchestration. The music is not abstract. Working with abstract AI generations forced me to adopt a different view on orchestration, where I had to dig into the material to find the best or most interesting way of showcasing it, rather than presenting it entirely as originally conceived. Thus, the relationship between material and function turned on its head: where I might normally create material with a certain function in mind, here I was discovering a function through closely considering the material. Devoting more time to understanding the material let me make some interesting orchestrational decisions throughout the piece that I might not otherwise have done (such as pairing two unusual instruments, or giving material only to the lowest instruments in a family, etc).

4.2.7 Musical Time

In *Silicon Mind*, I intended to take my interest in algorithmic time in a different direction to my previous works, such as *Three Entistatios III* and *Chromodynamics* which resulted in accelerating cyclic forms. Somewhat more speculatively, I also wanted to investigate what music of the future might sound like if an AI wrote music intended for other computers, not humans.

According to the definition of algorithmic time discussed in Chapter 1, time 'begins' with the first step and 'ends' when the algorithm is complete. It does not necessarily matter how long, in actual time, these steps take. This is an idea that chimes with some orchestral music. I imagined that a sonata form or the four movements of a traditional symphony could be considered a type of algorithmic time, where it is at least as informative to understand the relationship of the internal sections that unfold in a specific order than it is to count how many seconds have passed in actual time. I composed *Silicon Mind* in a warped sonata form partly to realise this connection between classical form and algorithmic or step-based time.

This kind of step-based time exists in any algorithm, but *Silicon Mind* also has major elements of musical time manipulation derived specifically from AI algorithms. AI algorithms learn from audio data like WAV files, or from symbolic data like MIDI files. Whether audio or MIDI, for AI training purposes this data can be transformed into an image, such as a spectrograph or a MIDI roll (Carykh 2017)[13].

Images do not exist in time – they are static. It's only when we tell the machine learning algorithm to play that image from left to right that the dimension of time suddenly originates. But if a machine was creating music for itself, in a theoretical future where machines exist that enjoy listening to music for its own sake, musical time probably wouldn't need to work in the way we hear it. The image-music could be enjoyed all at once, top-to-bottom, right-to-left or the traditional start-to-end. I wondered if this might be an equally valid way of hearing AI-generated music, even if it made less intuitive sense to a listener. This was a thought that I slowly developed through working with many types of algorithms that treat the 'problem' of time in different ways. This idea of AI algorithmic time builds on previous composers' work using technology-based time and form in

[13] Encoding music into another language, in this case visual, is not only theoretical – already in this commentary I have described musical data encoded into natural language data several times, and there are several research projects ongoing that are examining the utility of image-music translation (i.e., Oxford AI Society at the 2022 Sound of Contagion Workshop).

some way, such as Cassandra Miller's *Guide* (2013) and Sciarrino's *Efebo con radio* (1981).

To enact this in *Silicon Mind*, I created several axes of reflection across the piece (Figure 39). On either side of these axes, the same music is heard both forwards and backwards. This is not only retrograding rhythms and pitches, but also the timbre, decay, and attack of the sound. I imagined reading a spectrograph backwards (right-to-left) and therefore reversing the entire sound[14]. Often the entire orchestra is not in reverse, but rather some instruments flip at an axis of reflection, while others continue moving through the warped sonata form[15].

A major axis of reflection is in and around Bar 267. In this case, the reversed music is heard first before being played in its 'original form' after the axis in Bar 267. From Bar 277, the brass perform accented stab notes, which have already been heard in reverse from Bar 225. Similarly, from Bar 267 a solo violin performs a descending figuration derived from the first subject material (Bar 1) and percussion material (from Bar 228). Leading to Bar 267, this material is heard in reverse.

Realising this with only the physical instruments of the orchestra presented an enjoyable challenge. Many instruments can emulate a near-enough reversed sound by simply starting quietly and cutting off any resonance at the end of the note (e.g., brass from Bar 83). Some required more thought; reversing the sound of the vibraphone, for example, requires the percussionist to first bow the note to produce a sustained note, before

[14] This effect is used, for example, in Bars 76-89 in the brass. A harmonic progression is heard from Bars 76-81 before reaching an axis of reflection, specifically the F# in the 2nd Trumpet in Bar 82. Following this axis, the same music is heard in reverse, including reversing the timbre of the trumpets and trombones, who are using Harmon mutes to morph each individual note.

[15] This can be heard in Bars 149-171, where there is an axis of reflection after Bar 160. The single-reed wind instruments, 2nd violin, celli, and basses perform the music leading to Bar 160 in reverse, including emulating the reverberation of the accented note in Bar 160, while the rest of the ensemble continues to move towards a repeat of the entire exposition. Note that due to a metric modulation in Bar 160, the reversed semiquavers have been transformed to sextuplet semiquavers to keep them the same absolute length.

striking and damping the note with a mallet (e.g.., Bars 86-7). After a workshop day experimenting with the orchestra's principal percussionist, we found that many percussion instruments do not produce a sound comparable to their own post-strike decay when bowed. For certain instruments such as gongs (which produce a very different timbre when bowed to struck), I therefore emulated decay through other instruments in the orchestra (Figure 41).

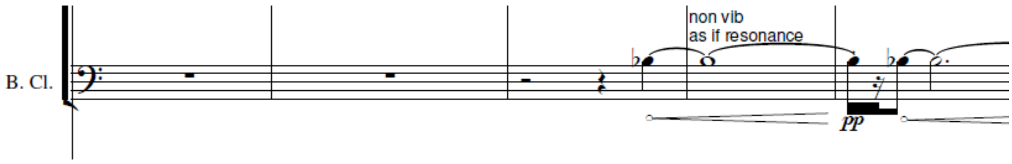

Figure 41: Example of emulating decay for percussion instruments. In this case the bass clarinet is used to provide resonance for tuned gongs, when the tuned gong material is intended to be heard 'in reverse'

Reversal of material is also heard on a larger, structural level. I realised that by placing an axis of reflection in the middle of the work but also by continuing some instruments moving forwards through the form, the exposition (reversed) would be heard at the same time as the recapitulation (forwards). By composing the first and second subjects such that they would harmonise when one was forwards and the other backwards, this seemed to me an interesting take on the traditional idea of transposing one subject into the tonic (Figure 39).

The same effect was also used for a particular wide scale used in all three movements of *Silicon* (Silicon Scale). This scale becomes sharper as it descends and is first seen in Bar 100. Because it becomes sharper every four notes, it does not repeat at the octave and requires the entire range of the piano to transpose back to its original form. I liked its effect when combined with a static G minor scale. This Silicon Scale and G minor combination bookends the development. In Bars 187-200, the wide scale is heard normally, and G minor is heard in reverse. In Bars 312-327 the Silicon Scale is heard in reverse and G minor normally.

The final way I wanted to explore a kind of AI-algorithmic form was through an idea of branching narratives. 'MuseNet' can be instructed to create any number of responses to a prompt, which will all be created simultaneously, and each will be different. Several times while composing, I orchestrated and included one 'MuseNet' generation before rewinding back to the start of that generation to use another, creating a sonification of constant progress through many iterations of the same task. This was inspired by a conversation I had with 'MuseNet' programmer Christine Payne, who mentioned she was interested in including such a 'possibility-space' in a future algorithm. I used a ratchet and kick drum to signify sonically when the music was reverting back to an earlier point of divergence[16].

4.2.8 Microtones and Structuralism

When developing my thoughts on musical structuralism through consideration of machine learning data augmentation techniques (Chapter 2), it struck me that (taken on its own terms) data augmentation had not gone far enough. To summarise, I had been surprised by the way that computer scientists process musical data for machine learning, by applying transformations such as all-intervalic transposition and time stretching. This implied that the music emerged from relationships between data points, and that the music would be fundamentally the same if you changed any actual information about the music, so long as the relationship between data points was kept proportionally the same.

If this is the implication, why stop at the piano's 12 tones of transposition? Why not also transpose the music microtonally (to a potentially infinite

[16] For example there is a ratchet reversion at Bar 11 which reverts to Bar 1 and at Bar 41 which reverts to Bar 37.

number of transpositions), if only the relationship between pitches and rhythms matters? I wanted to include microtonal transposition of material in *Silicon Mind* as a kind of rejection of this thought process, but also to explore where the thought might lead me (e.g., Figure 42).

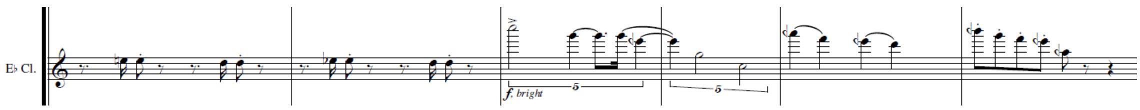

Figure 42: First subject of 'Silicon Mind' altered with microtones during the exposition repeat

Microtonal transposition fitted well with my intention to create something uncanny in the middle of *Silicon Mind*. As the material from 'MuseNet' became more and more polarised, it slips into a microtonal landscape until by Bar 205 the orchestra is playing functional microtonal harmony. When the percussion material begins at Bar 228, it is intended to sound 'out of tune' (even though it is in regular tuning), adding to an 'uncanny valley' feeling in the music (Figure 43).

Figure 43: Percussion enters on top of microtonal chords in strings (which are alternating between F quarter-sharp and C quarter-sharp harmonies)

4.3 Silicon Body

One of the biggest concerns raised by AI might be that of authenticity. In recent years we have become familiar with AI's capacity for creating believable fakes. This technology is used to automatically generate stories that resemble human-written news and by social media giants to encourage engagement, with the dissemination and promotion of fake news stories as known by-product (Wang et al 2018). It is now a regular occurrence to see AI algorithms used to create fake videos showing public figures in unfavourable light (Botha & Pieterse 2020) and it has also been used in movies to allow deceased actors to appear in new releases (e.g. Peter Cushing and Carrie Fisher in *Rogue One: A Star Wars Story)* or to de-age live ones (Sargeant 2017). This kind of technology is often colloquially called deepfake, but the technical term for this field is Style Transfer (see Appendix 2).

Accordingly, we are now becoming used to questioning the provenance of believable-looking sources in a way that we were not even at the turn of this century. Authenticity is becoming more important as an issue, and discussion of authenticity might not be limited to only identifying whether something is AI- or human-generated. For example, this year (2022) has seen the release of several generative visual art algorithms (e.g., Dall-E 2; Wombo; Dall-E Mini) that have spawned multiple news articles on authenticity (i.e., Schreiner 2022). Real is not the same as authentic – and this is especially clear in the creative fields. An AI might generate *real* music in the style of Mozart (as discussed), but this music might not feel *authentic* to all listeners. Authenticity is a much more subjective question. Here, the orchestra, and classical music more generally, can offer a perspective.

Classical musicians are familiar with questions of authenticity. Discussions and disagreements emerging around, for example, performing Bach on the modern piano (Edidin 1998), using vibrato during 18th century symphonies

(Norrington 2004), or casting singers of colour to sing operatic roles representing minority groups (see André et al 2012) can be viewed, at least partially, as questions of authenticity. Though this by no means exclusive to this one genre, it remains true that classical musicians are plugged-in to whether music feels authentic, in addition to what it sounds or looks like on the surface.

This leads to the questions that at the foundation of the second orchestral movement, *Silicon Body.* What exactly is fake music? And does fake or inauthentic music become any more authentic when performed by an orchestra – by real people? Perhaps most importantly, I wanted to hear what this deepfake technology sounds like. I wanted to embed an instrument that uses AI deepfake technology within the orchestra, to be played by an orchestral musician, as a kind of model for one way orchestras might be constituted in the age of AI.

One research paper that particularly interested me showcasing deepfake technology is called 'Everybody Dance Now' (Chan et al 2019). It demonstrates taking a video of a dancer (Source), an image of a second person (Target), and the use of AI to make the Target appear to move like the Source. To do this, it strips the Source video down to a basic set of moving points and lines, abstractly representing the human body. With this distilled from the Source, the AI then rebuilds the video, this time with the Target fleshing out the skeletal nodes. I found the way that computer vision 'sees' people fundamentally differently to how we see people fascinating – and a perhaps a little unnerving. An answer to what fake music might sound like lay, for me, in the relationship between the surface – the Target – and the hidden layers – the Source.

Silicon Body has a Source, a layer of music that sits underneath and drives the whole piece. This skeletal musical framework is made up of three alternating patterns of pitches and rhythms. These can be seen in Appendix Score 2 (an extract of the Source part), and are as follows:

- Descending scales beginning on G [Appendix Score 2: Letters A and E]
- Ascending and Descending Silicon Scale [Letters C and G]
- Descending and ascending fourths and fifths [Letters B, D, F, and H]

The Source layer moves through and mutates these ideas in turn, with the aim of realising G minor at the same time as the Silicon Scale. At this point, its logical argument finishes, and the movement finishes soon after.

The Source is performed by a digital instrument called DDSP (Differential Digital Signal Processing) developed by Google Magenta (Engel et al 2020). DDSP is a Style Transfer instrument that works in a similar way to the earlier dancer example, except that the Source and Target are audio-based rather than video-based. We can play any sound into DDSP and instruct it to transfer that sound's harmonic content into any other timbre using AI. It's also possible to push the instrument outside of its intended comfort zone to create exciting new timbres through AI. I worked with Magenta ahead of DDSP's release, providing recordings of my performer colleagues that the model could learn from and imitate during *Silicon Body*. The orchestral keyboardist performs on a synthesiser, which is linked to DDSP through Ableton Live (Figure 44).

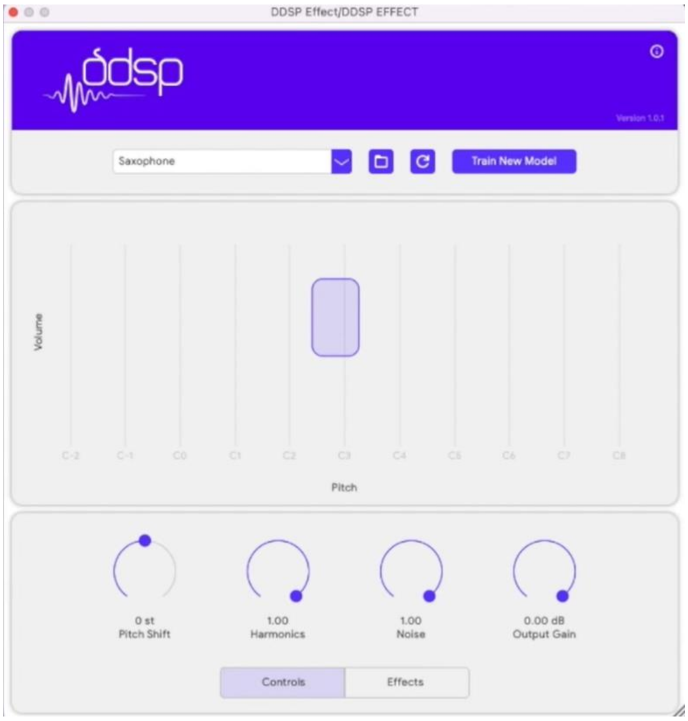

Figure 44: DDSP Interface on Ableton Live. User can select model (in this case saxophone), pitch range, and relative volume of harmonic and noise content

On top of this Source are superimposed three Target styles of music that are performed by the orchestral instruments. Inspired by 'Everybody Dance Now', these three styles are based on different types of dance music – big band jazz, electronic techno, and folk. Continuing the idea of reference to traditional orchestral music, this also makes *Silicon's* second movement a kind of dance movement, to follow the first movement's sonata form references. I composed the jazz and techno styles in their entirety, in short score.

For the big band jazz style, I modelled it on music I have personally played in big bands (such as music by Gordon Goodwin, Glenn Miller, and Michael Giacchino) and on music inspired by New Orleans Second Line drumming, such as *Quartermaster* by Snarky Puppy. This resulted in a verse-chorus-solos structure overlayed atop a second line-style drum groove (Figure 45).

Figure 45: Jazz Style Short Score

To compose the techno style, I analysed some two dozen pieces in this style recommended to me by a DJ colleague, including pieces such as *Clipper* by Autechre, *Thinking of You* by Herbert, and *C45p* by Helena Hauff. With these as my blueprint, I compiled a grid of different materials that would fade in and out across the course of the techno style (Figure 46). To me, one of the most interesting elements of this genre is the timbral complexity of any given sound, allowing it to be repeated many times in a pattern without becoming dull. Therefore, the techno style for *Silicon Body* was focussed more on timbre and tone, while the jazz style focussed more on harmony.

Figure 46: Techno Style Material Grid

After composing and discarding a short score for the folk-dance style, I decided to utilise Folk-RNN for its material (see Appendix 1). It seemed to be fitting to use AI to generate some of the music. I returned to an earlier piece, *Three AI Folk Songs*, to orchestrate instead (Appendix 1 and Figure 47). In that earlier work, the Folk-RNN material was presented more-or-less exactly as the AI wrote it, but in *Silicon Body* I wanted to push it into uncanny territory. I did this through, for example, applying microtonal glissandi to the melody (Figure 48).

Figure 47: Folk-RNN Original

Figure 48: Folk-RNN orchestrated. Begins Bar 26 of Figure 4. Note it has been transposed down to fit the Source harmonic content

Silicon Body essentially consists of each of these three Target dance styles occurring simultaneously, but only one is usually heard at any given moment. The surface of the music cuts between the three styles at an ever-increasing rate (Figure 49), rotating through them faster and faster until the music reaches a breaking point. At this point, the same point that the Source concludes its logical argument, the Source is revealed on its own and the piece then ends. Once I had determined where the cuts would be, I knew where the three styles would occur, and which sections of the short scores would be used in the piece proper. I then orchestrated these sections of short score into full score. Each style has their own set of instruments that do not usually overlap.

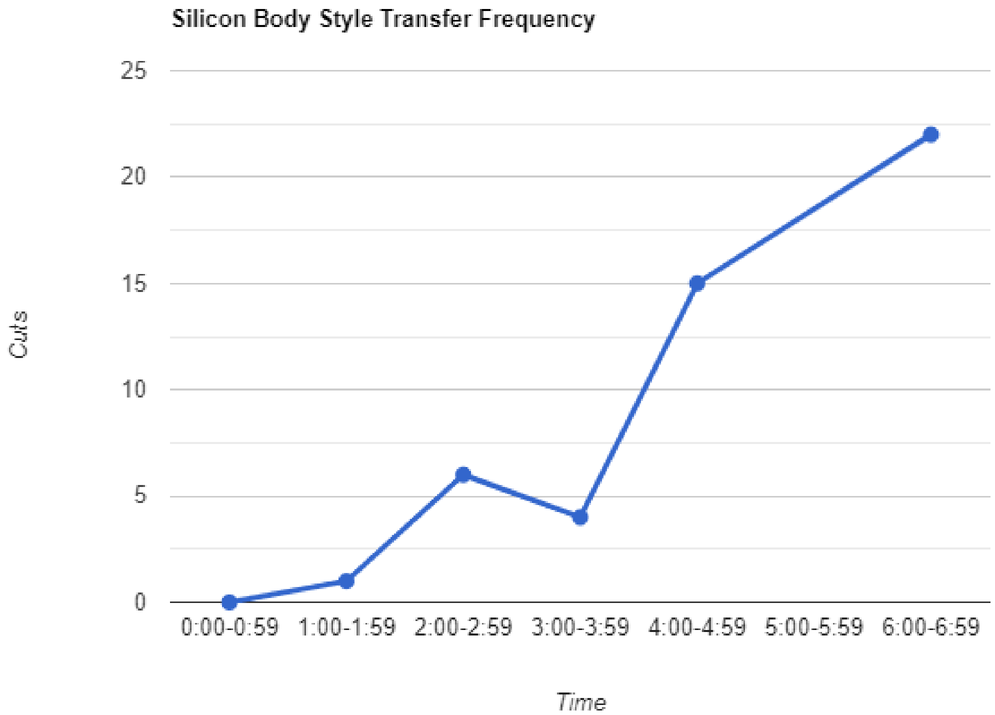

Figure 49: Number of cuts between styles (or to no style) in each minute of Silicon Body. Note that the styles overlap during 5:00-5:59 and after 7:00, making tracking discrete cuts impractical

It was my intention that the piece would slowly feel more inauthentic as it became clear that the surface layer of the music was completely governed by a more abstract and alien Source layer. The Source layer determines most musical properties of the Target layer, including tempo, tonal area, and rhythm, forcing the Target styles to be transposed, stretched, or crushed according to the Source layer logic (Figure 50).

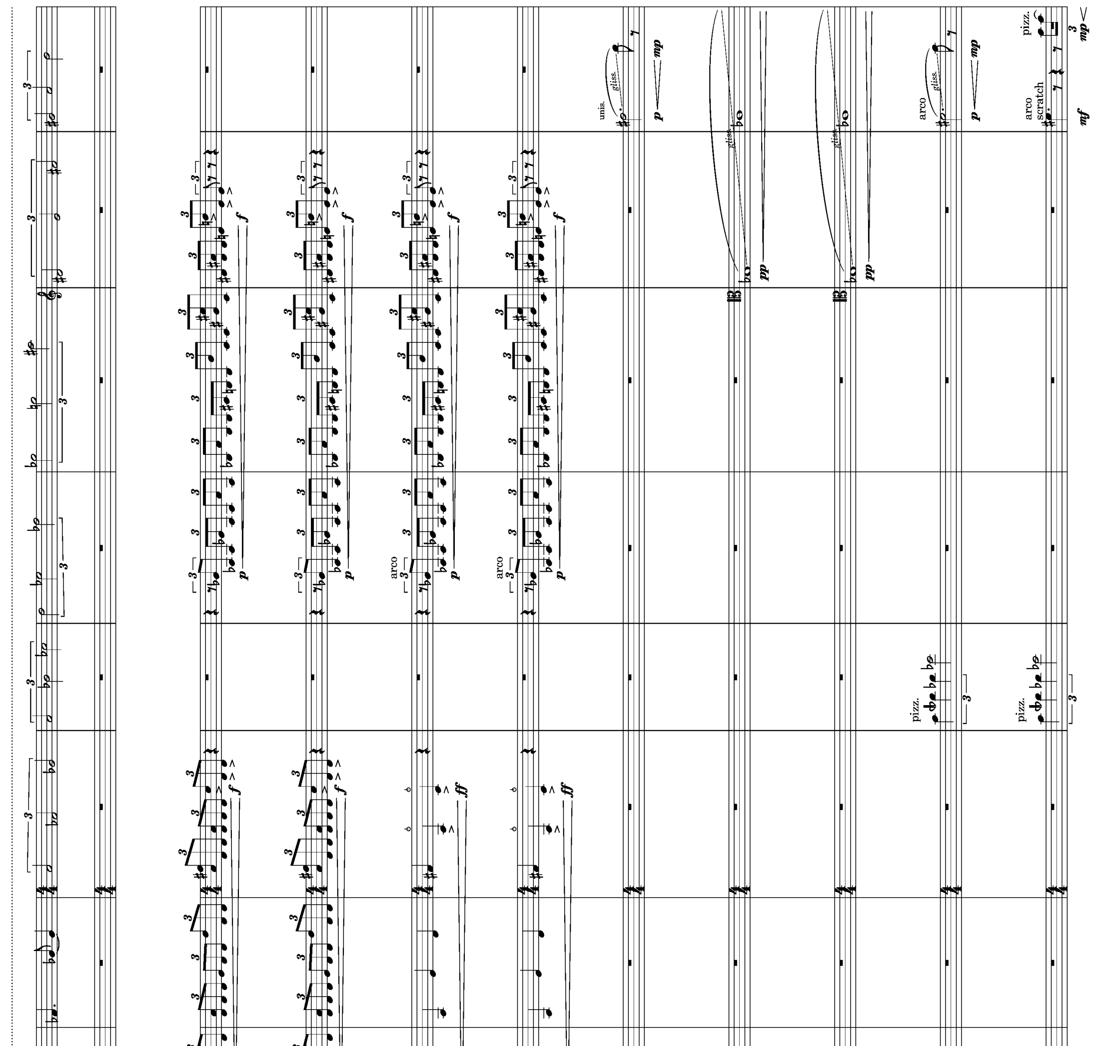

Figure 50: Example of orchestral material (Target) transposing itself to match harmonic content of DDSP (synthesiser)

The music is meant to sound fun, uncanny, and sinister, reflecting the many uses of deepfake technology. It is also intended to evoke another type of AI algorithm that atomises attention spans – those that govern social media websites (Hao 2021).

4.4 Silicon Soul

The third movement, *Silicon Soul*, was intended to probe two areas I felt I had not yet given enough attention. The first was technical: I wanted to better incorporate audio-generative AI into my work, and I also wanted to test the limits of the PRiSM-SampleRNN AI algorithm (Melen 2020) that had been developed over the pandemic by providing it, via Christopher Melen (the PRiSM Software Engineer), with challenges regarding dataset size and audio quality.

Until this point, most testing done in-house with SampleRNN utilised a training dataset (from which the AI learns) of approximately 0.5 to 10 hours in length. My collaboration with OpenAI on MuseNet and use of other large models had made me aware of the potential benefits of using a much larger dataset for general-purpose training, and I wanted to test this with SampleRNN. The BBC Philharmonic provided me with access to their archive of broadcasted concerts, which I turned into a dataset lasting approximate 2000 hours.

Additionally, SampleRNN had mostly been used to generate relatively low-quality audio, at 16,000kHz. This is because a lower sample rate provides a host of benefits, primarily making testing, training, and generating audio much faster. The model for *Silicon Soul* was trained to produce audio at 44,100kHz (CD quality) instead.

Both challenges vastly increased the amount of time that SampleRNN took to train, but when the training was complete, I was left with an AI model that felt far superior for incorporating into my creative process than previous tests with this AI (see Chapter 2). More specifically, I had five different models, representing different stages of the training process. Each model had its own ‘sound’ – its own take on how to imitate the BBC Philharmonic (Figure 51).

AI Model Name	Notes
BBC-Full Epoch 1	Very subtle at lower temperatures, possibly useful as a morphing background texture. Timbrally sometimes half sine-wave, half orchestral sounds. Recognisable progression towards very symphonic (Straussian) brass sounds as temperature is increased. Loud strings, little woodwind.
BBC-Full Epoch 2	Fuzzy and volatile. Very low audio quality, with booming and clipping bass. Higher temperatures generally slightly higher quality. Unlikely to use Epoch 2 generations in this piece.
BBC-Full Epoch 3	Lower temperature (0.9-0.95) has beautiful and calm textures, quite ethereal. 0.975 temperature retains this quality but with occasional flashes of recognisable orchestral activity. Some generations transform into or out of applause.
BBC-Full Epoch 4	Very slow moving and languid. Mysterious in places – good for supporting orchestra or collaging on top of itself. Very high quality audio.
BBC-Full Epoch 5	Very high quality audio again. Epoch 5 generations are in motion, exciting and engaging. At lower temperatures the bass is fuzzy but this goes away from 0.975 onwards. 0.99 temperature generations are very exciting and symphonic – they could be actual recordings. Use for climax of electronics part.
Presenter-Only	Lots of useful material here. Plenty of presenters speaking, they speak in what sounds like a garbled made-up language (Sample-RNN cannot learn semantics, only the spectral profile of speech). Lots of applause and tuning notes, sometimes blending with presenter's voice. After some 40 generations from this dataset, I have never heard the AI imitate a female presenter voice. Interesting comment on the overall trends of BBC Radio 3 archive recordings.

Figure 51: My original notes on each Sample-RNN model (Full orchestral dataset and presenter-only)

My dataset was not only music, but entire broadcasts from BBC Radio 3. This is because the BBC Philharmonic is a broadcast orchestra, whose remit is to provide content for radio. In practice, this meant that my dataset had other radio-like material within it, such as audiences applauding, presenters introducing music, and the orchestra tuning in the background. I realised that if I wanted to make an AI respond specifically to the BBC Philharmonic, I should also include these sounds – they are, in my opinion, part of the DNA of a radio broadcast orchestra. I created a separate dataset of just these non-musical Radio 3 sounds (around 5 hours) and trained a separate SampleRNN model on it (Figure 51).

I used the results of these six AI models to create an electronic accompaniment to the orchestra. I wanted the two to be roughly in sync without requiring the conductor to be beholden to a metronome, so I divided this accompaniment into several smaller tracks which are triggered by sample pads played by percussionists, allowing each track to begin at the correct time. The electronic accompaniment moves from the non-musical audio results through to volatile and dynamic imitation of full orchestra.

While creating this electronics accompaniment, I returned to the idea of collaging I had developed in earlier pieces. Sample-RNN audio generations are monophonic, and if the same settings are used for two generations, they will create two very similar pieces of audio (but never identical). I decided to layer several similar generations atop one another to create a shifting, stereo-like effect (Figure 52). This layering of material from the same AI model was more effective for audio generations than symbolic ones, and I feel that I achieved a more successful 'immersive' effect than similar collaging attempts using MuseNet in *Three Entistatios* and *Alter.*

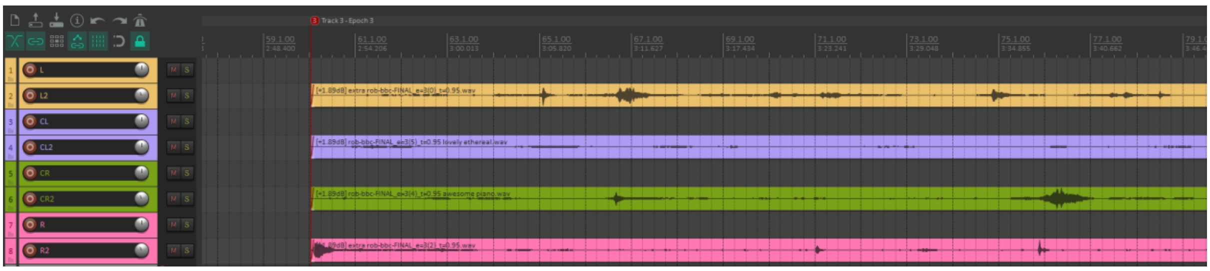

Figure 52: Example of Collaging Sample-RNN Material to Create Electronics for Silicon Soul

I wanted to showcase the AI as it actually sounds, so beyond layering (which, in my view, assists in getting a general feel for ‘how it sounds’) and choosing which to include in the piece, I did not edit the AI generations. At several points, the orchestra makes room for the electronics by playing very quiet and simple material, or not playing at all.

The second area I wanted to explore with *Silicon Soul* was aesthetic: I wanted to explore my third duality, which I called System and Secret. To do this, I imagined a ‘perfect’ AI algorithm as a thought-experiment. The thought-experiment algorithm has no artefacts, and it can achieve whatever musical task we set it. It can analyse any amount of data, unrestricted by hardware limitations, and can produce new data (i.e., music) trivially quickly. It can produce sound indistinguishable from human musicians in any genre, historical period, or ensemble. It can even produce entirely new music by combining existing music in novel ways or identifying gaps in its dataset that have never been exploited. But is it music, or does it only *sound* like music?

Would people accept this music, or do we require some kind of secret ingredient in order to feel a genuine connection with art? We don’t know the overall answer to this question, or even if it can be answered, in this specific instance because AI has not reached the fluency of the thought-experiment, but it’s reasonable to imagine it will. And if I take the view that there is more to music than computer data can communicate, what is that secret? Does it exist inherently within the music, or can this secret be imagined or imposed by the audience? Will AI research, in its pursual of a systematic, mathematical, and function-based understanding of the world, help us understand what the secret of music is?

This and related questions are already under active consideration from a wide range of artists and academics who have influenced my thought. Federico Campagna argues that embracing a worldview he terms “magic”, informed by elements of spiritualism, mysticism, and religion, can help alleviate the difficulties, both personal and social, inherent in a worldview reliant on data (Campagna 2018). Similarly, the authors of the *Atlas of Anomalous AI* explicitly state their aim to 're-mythologise AI in a way that reveals the roots of the technological project in spiritual practices, institutions and frameworks' (Vickers & Allado-McDowell 2021). George E. Lewis describes a view of improvisation as 'something essential, fundamental to the human spirit', before going on to assert that attempting to teach computers to improvise 'can teach us how to live in a world marked by agency, indeterminacy, analysis of conditions, and the apparent ineffability of choice' (Lewis 2018).

I set out to provide one response to this question by examining it through the lens of orchestral music. I wondered why audiences still go to see the orchestra today. As the pandemic has shown, it is perfectly possible to livestream performances to tune into from home, and there are also sample libraries that allow us to emulate the orchestral sound without needing any humans at all. What is its secret that compels people to physically come and watch humans make these sounds live?

For me personally (I am not trying to generalise), it is in understanding an orchestral performance not primarily as an act of creating sound, but rather as an act of community shared between musicians. For the third movement of *Silicon,* I wanted to experiment with including AI inside such a framework.

To do this, I intended to make the AI personal to *that* orchestra, as described above through specific use of dataset. In this instance AI is used as a tool to increase the personalisation and site-specific nature of a piece, rather than as a tool to make general rules about music. It is in service of defining what the nature of *this* ensemble is, and in using it I was challenged to make decisions about how to treat the similarities and differences in sound between the physical orchestra and its AI doppelganger.

Drawing on my earlier thoughts on musical structuralism, the orchestral part was intended to be non-transposable. I wanted each note to make sense only on the instrument it was written, and in that way for the material to be tied to the essence of each sound, rather than the relationship between sounds. To achieve this, I expanded the Silicon Scale to cover the entire range of the piano, then assigned each instrument one note from this wide scale (with only a couple of exceptions). Each instrument was given a note that seemed 'natural' to that instrument, such as open strings, natural harmonics, notes that sat firmly within the instrument's ideal playing range, or those with special significance (such as the oboe's A natural 'tuning' note) (Figure 53).

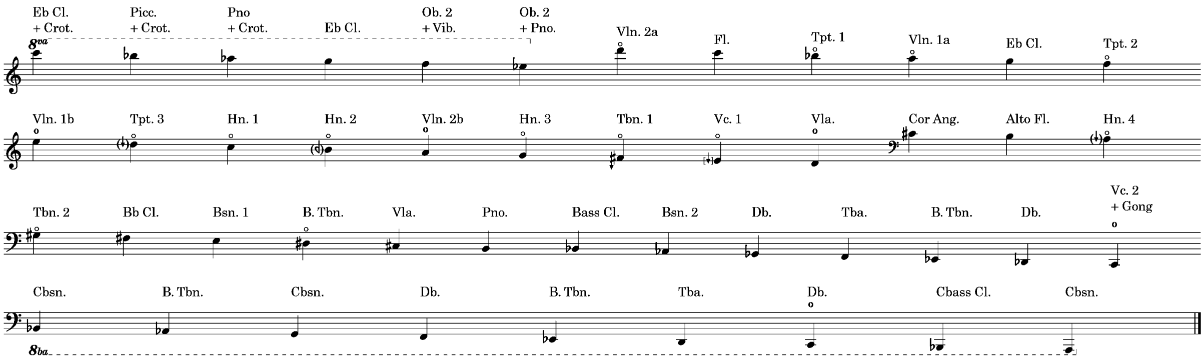

Figure 53: Silicon Scale notes assigned to orchestral instruments. Any brass harmonics are natural (no valves). String notes are either open strings or natural harmonics. Eb Clarinet notes derive from the initial melody in Silicon Mind

Expanding this idea of 'natural' sounds, I searched for non-pitched sounds that could only be made by specific instruments in the orchestra. My approach to extended techniques in my work usually comes from a place of pushing what is 'normal' to an extreme, rather than totally reinventing performing conventions on an instrument (see arrays in Chapter 3). Working with some performer colleagues, I found a series of non-pitched sounds in the strings, percussion, and brass with which to begin the piece that accompanied the fuzzy quality of the Sample-RNN towards the start of the electronics track. Therefore, the orchestra and the AI are both acting like 'shells' of a regular orchestral performance at various points in the piece: the orchestra through retaining only extended techniques, and the AI through acting as a strange audible mirror.

For the majority of the piece, each instrument only plays its assigned note or extended technique. The rule is broken when the orchestra imitates material generated by Sample-RNN (Figures 54 and 55), and at the end of the piece where the descending Silicon Scale is played by several instruments.

Figure 54: Orchestra imitating Sample-RNN 1

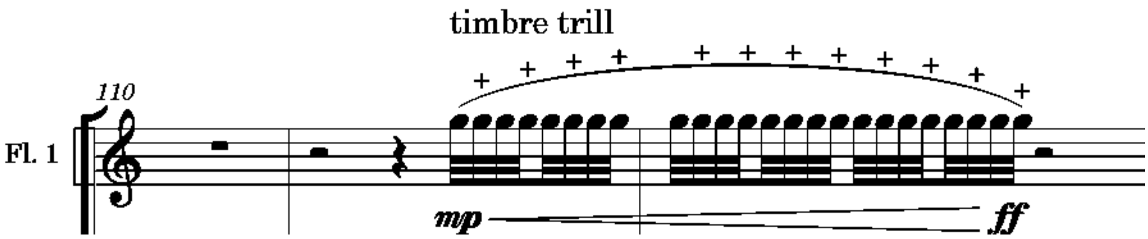

Figure 55: Orchestra imitating Sample-RNN 2

To finish the piece, I wanted the orchestra and the AI to play completely shared material. I decided to return to the combination of G minor and Silicon Scale to do this. Sample-RNN always generates its material autonomously and cannot be instructed exactly what to play. For this section, I used another audio-generative algorithm called RAVE, released by IRCAM. RAVE is similar to DDSP in that it can learn how to imitate a dataset then apply that sound to pitches and rhythms given by a user. I trained RAVE on the same BBC Philharmonic dataset as my SampleRNN models and wrote a part for it in the score. RAVE then realised this part as though it were the BBC Philharmonic.

5: Conclusion: What is AI?

This commentary has documented several ways I have used AI within my creative process and identified areas of aesthetic interest that have arisen across my artistic work as a result of working with AI. There is, however, an elephant in the room it has not addressed: what is AI in my practice?

When I began this PhD, I tended to think of AI algorithms as a tool in my arsenal, to join the other compositional tools I already used on a regular basis such as harmonic rotation, isorhythmic practices, and large-scale formal arguments. In some sense, AI is a tool. DDSP, for example, could be described as a tool – it did what I wanted it to do in *Silicon Body*. But in other ways, AI is nothing like a tool. A hammer is a tool, but no matter how long you leave a hammer on the counter, it will not learn to build furniture. AI can teach itself to carry out tasks – as shown clearly through MuseNet, Sample-RNN, and the reinforcement learning algorithms used in *Rose Green*. DDSP, too, is autonomous – the excitement of this algorithm, for me, came from the unexpected things it learned from instrumental datasets, and the uncanniness of its attempts at perfect sonic recreation. I have never before seen a tool that teaches itself rules to carry out its appointed task, so I don't think this is the correct terminology. At best, in my practice, AI has occasionally acted 'tool-like'.

For a period during my PhD, I thought of AI as a collaborator. This seemed at first somewhat appropriate when there was a clear delineation of contribution to a creative process. Using AI to generate new texts and then setting them to music entirely myself is, on the surface, similar to collaborating with a poet or librettist on a piece of music. However, I ultimately discarded this term because of its anthropomorphic connotations. AI does not understand the decisions it is making, it only understands specific mathematical functions and goals. It is not capable of conversation, discussion, disagreement, or development of an idea in tandem with another person. That is not to say AI will never reach the level of collaborator, but that it has not done so for me yet. In other fields, it is much closer. Large natural language models are capable of some or all of the qualities I just described. Music AI is just not there yet. To me, collaborators are always the programmers who code the AI. It is with them that I collaborate, and the results of our collaborations are AI models.

I could consider AI in my practice, then, to be something that *makes decisions*, in which case I could call it an agent. At time of writing, this is probably the most accurate single word to describe what AI is to me. It acts as an agent within my compositional process, making decisions that are informed by the rules it has learned, which equally inform my global creative process for that work. This is accurate in an analysis of an individual piece – the analysis of *Alter* might be the best example for AI acting as an agent within a wider creative context. This does not exactly get around the issue of AI fundamentally not understanding the decisions it is making, though it does help to de-anthropomorphise the way in which I might refer to it. A bigger issue with using this term on its own is that it does not encompass the myriad ways in which AI has impacted my compositional work in the abstract, away from specific decisions made during the creation of any given piece.

As exemplified in Chapter 3 and 4, my interest in AI goes beyond just the notes, rhythms, and sounds it can generate. It extends to the kind of music we might make in a world dominated by AI, and how we might respond to AI that impacts all areas of society, both positively and negatively. My work responds to AI as an idea just as often as it uses literal machine learning as part of the creative process. Calling AI within my creative process an 'idea', however, encounters in relief the same problem as calling it an 'agent'. There are plenty of inspiring artists who use AI as a muse without ever using, or even having any knowledge of, how machine learning actually works, and the real-life creative opportunities machine learning offers. For me, the idea of AI should not come at the expense of the genuine ground-breaking opportunities this technology brings.

In summation, there is no single word that describes what AI is in relation to my work. It simultaneously acts tool-like, agent-like, and idea-like. It cannot be defined as simply one thing for me. In this way, I think it is like the Internet, or film, or the printing press. These are technologies that changed the underlying way modes through which we communicate with one another, the media through which artists work, the social environment in which they operate, and ultimately the way people think. AI is causing these fundamental seismic shifts and, from my perspective, is only becoming more important as the technology becomes more advanced and more accessible.

Taking my practice forwards, then, will mean an even closer marriage of learning to use technology in my work with a deeper understanding of what it means to be making music in a world of increasingly advanced technology. My compositional research will continue to use AI, which has found a secure and genuinely useful home within my work. I plan to broaden my scope to other advanced technologies, such as extended reality, wearable technologies, and other technologies that blur digital spaces with physical ones. This PhD has focussed heavily on orchestral music, and the incorporation of advanced technology into large-scale forces (or large-scale forces into new technology) will continue to be an area of active interest. However, I am also aiming to explore what 'chamber music' might mean in the age of advanced technology.

While some of the AI algorithms I have written about in this commentary may become outdated, perhaps quite quickly in some cases, I hope that the thoughts and techniques I have developed through using them will remain relevant and provide a basis for my compositional development in the future.

Glossary

Audio-Generative or Audio-Based

An algorithm that produces audio files (such as WAV or MP3) as its generations, rather than text, MIDI, or any other form of data. These types of algorithms tend to learn from and generate music on a sample-by-sample basis.

Dataset (also: Corpus)

The data that an algorithm learns from. An audio-generative model might learn from a dataset of WAV audio files, and a symbolic-generative model might learn from a dataset of MIDI transcriptions.

Deep Learning

The subset of machine learning that uses neural networks.

Encoded

To encode data is to transform it from one form of representation to another. In this portfolio, music and text are often encoded and decoded into one another.

Fine-Tune

To fine-tune a model is to first train a model on a very large amount of data, so that it learns general rules, then to provide it with a small dataset that the user wants the model to specifically emulate. This allows the model to learn more rules than it could from only the small dataset, but avoids the model producing unwanted data *from* the large dataset.

GAN (Generative Adversarial Network)

From Techtarget.com:

"A generative adversarial network (GAN) is a machine learning (ML) model in which two neural networks compete with each other to become more accurate in their predictions. GANs typically run unsupervised and use a cooperative zero-sum game framework to learn.

The two neural networks that make up a GAN are referred to as the generator and the discriminator. The goal of the generator is to artificially manufacture outputs that could easily be mistaken for real data. The goal of the discriminator is to identify which outputs it receives have been artificially created.

As the feedback loop between the adversarial networks continues, the generator will begin to produce higher-quality output and the discriminator will become better at flagging data that has been artificially created."

General-Purpose

An AI that has been trained on a very large dataset and is therefore capable of undertaking many different tasks. For example, OpenAI's 'GPT-3' algorithm or Google's 'BERT'.

Generations

The result of an AI model; what an AI model generates.

Generative AI

The class of AI that is concerned with generating new data, as opposed to (for example) analysing or classifying data.

Image-Generative

A type of generative AI that produces visual images.

[Large] Language Models

See NLP

Model

The model is the result of an algorithm's training. An AI algorithm creates a model, and it is generally this model that the user interacts with (as opposed to the algorithm itself).

NLP (Natural Language Processing)

The area of AI that deals with learning and understanding written language.

'Other AI'

My term for algorithms in this commentary that produce neither audio nor symbolic music data.

Prompt

An AI model can be given a prompt, which is some data the user wants the AI to continue. For example, I can provide 'MuseNet' with a prompt of four bars of music, or I can provide 'GPT-2' with a prompt of the beginning of a sentence. Using prompts allows the user to direct the content of the AI somewhat.

Reinforcement Learning

An area of AI research that is separate to deep learning. It concerns algorithms who learn to undertake a specific task through goal-oriented trial and error, without a dataset.

RNN (Recurrent Neural Network)

A type of AI algorithm used widely in text and music generation. RNN architectures are used for generating data that exists in time, or in sequence, such as music or language. In these cases, what is being written now depends on what has been written before.

LSTM-RNN (Long Short-Term Memory Recurrent Neural Network)

An architecture for an RNN algorithm that improves its long-term coherence. At the time, LSTM architectures for RNNs were one of the most promising areas of music- and text-generative AI research.

Style Transfer

See Appendix 2.

Symbolic-Generative

AI algorithms that generate representations, such as text or music notation. Usually contrasted, in this commentary, with audio-generative algorithms.

Temperature

A common parameter in generative algorithms, adjusting the temperature adjusts the 'riskiness' of an AI's decisions. A high temperature on a text generator, for example, might produce erratic and incoherent texts, while a low temperature might produce the same sentence repeatedly.

Text-Generative

AI algorithm that generates written text.

Token

Terminology used mainly by OpenAI to denote an arbitrary unit of information. Tokens are used in 'GPT-2' and 'MuseNet' to represent characters, words, notes, rests, and chords. Designing architectures based on tokens allows more flexibility within the network than some other methodologies.

Training

Training describes the process by which an AI learns rules from its dataset. This process involves the algorithm examining its dataset repeatedly while updating its internal statistical functions. Depending on the architecture of the algorithm and the type and size of the dataset, training can take a very long time.

Transformer

Type of AI architecture that superseded LSTM-RNNs for music and text generation around 2019. This is because it could generate coherent results over a longer timespan.

Vocal Synthesis

AI that synthesises the human voice. Often used for text-to-speech services.

Text-RNN

A recurrent neural network (RNN) that models language and generates text. In this case, it was on a word-by-word basis. That is, it learns to estimate which word is likely to follow any given word. This is distinct from a character-based RNN, which would make this statistical decision after each character. The advantage of a word-based RNN is that it will rarely produce nonsense words. We used an LSTM architecture for the text-RNN in *Alter* because of our familiarity with this style of model from knowledge of 'Clara' and other LSTM-RNNs used for modelling polyphonic music (e.g., Ycart & Benetos 2017).

References

AIVA. *AIVA - The AI composing emotional soundtrack music* (AIVA, retrieved April 4 2022), https://www.aiva.ai/

Alcubierre, M. (1994). The warp drive: hyper-fast travel within general relativity. *Classical and Quantum Gravity*, 11(5).

Alpern, A. (1995). *Techniques for Algorithmic Composition of Music*. Dissertation, Hampshire College. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.9364

Anders, T. (2018). Compositions Created with Constraint Programming. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 133–154). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.5

Anderson, J. (2002). Harmonic Practices in Oliver Knussen's Music since 1988: Part I. *Tempo,* (221), 2-14. doi:10.1017/S0040298200015618

André, N. A., Bryan, K. M., & Saylor, E. (2012). *Blackness in opera*. University of Illinois Press.

Bahuleyan, H. (2018). *Music Genre Classification using Machine Learning Techniques*. ArXiv. *arXiv:1804.01149 [cs.SD]*

Brock, A., Donahue, J., Simonyan, K. (2019). *Large Scale GAN Training for High Fidelity Natural Image Synthesis*. ArXiv. *arXiv:1809.11096 [cs.LG]*

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Openai, W. Z. (2016). *OpenAI Gym*. ArXiv. *arXiv:1606.01540 [cs.LG]*

Böck S., Krebs F., and Widmer G. (2015). Accurate Tempo Estimation Based on Recurrent Neural Networks and Resonating Comb Filters. *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR, Málaga, Spain*, pp. 625-631

Botha, J., & Pieterse, H. (2020). Fake News and Deepfakes: A Dangerous Threat for 21st Century Information Security. In B. K. Payne & H. Wu (Eds.), *ICCWS 2020 15th International Conference on Cyber Warfare and Security*, pp. 57–66.

Caillon, A., & Esling, P. (2021). “RAVE: A variational autoencoder for fast and high-quality neural audio synthesis.” ArXiv. *arXiv:2111.05011 [cs.LG]*

Campagna, F. (2018). *Technic and Magic*. Bloomsbury Academic. https://doi.org/10.5040/9781350044005

Carykh, July 5th 2017, *AI Evolves to Compose 3 Hours of Jazz!* [video], YouTube, https://youtu.be/nA3YOFUCn4U

Chan, C., Ginosar, S., Zhou, T., & Efros, A. A. (2019). Everybody Dance Now. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 5933–5942.

Collins, N. (2018). Origins of Algorithmic Thinking in Music. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 67–78). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.2

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. *AI and Society*, 1, pp. 1–25. https://doi.org/10.1007/S00146-021-01294-X/FIGURES/11

Cox, G., & Riis, M. (2018). (Micro)Politics of Algorithmic Music. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 603–626). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.22

Criado Pérez, C. (2020). *Invisible Women: Exposing Data Bias in a World Designed for Men*. Penguin.

Dastin, J. (2022). Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women. In K. Martin (Ed.), *Ethics of Data and Analytics* (pp. 296–299). Auerbach Publications. https://doi.org/10.1201/9781003278290-44

Dhariwal P., Jun H., Payne C., Wook Kim J., Radford A., Sutskever I. 2020. *Jukebox: A Generative Model for Music*. ArXiv. *arXiv:2005.00341v1 [eess.AS]*

Donahue, C., Mao, H. H., Li, Y. E., Cottrell, G. W., & Mcauley, J. (2019). LakhNES: IMPROVING MULTI-INSTRUMENTAL MUSIC GENERATION WITH CROSS-DOMAIN PRE-TRAINING. *20th International Society for Music Information Retrieval Conference*, pp. 685–783. https://chrisdonahue.com/LakhNES

Donne, Women In Music, *2019-2020 - Donne, Women in Music Research* (Donne 2020), https://donne-uk.org/2019-2020/

du Sautoy, M. (2016). *What We Cannot Know*. Fourth Estate.

du Sautoy, M. (2020). *The Creativity Code: Art and Innovation in the Age of AI*. Fourth Estate.

Edidin, A. (1998). Playing Bach His Way: Historical Authenticity, Personal Authenticity, and the Performance of Classical Music. *Journal of Aesthetic Education*, 32(4), pg 79. https://doi.org/10.2307/3333387

Einstein, A. (1920). *Relativity: The special and general theory*. Penguin.

Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., & Simonyan, K. (2017). Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders. *Proceedings of the 34th International Conference on Machine Learning*, pp. 1068–1077.

Engel, J., Hantrakul, L., Gu, C., & Roberts, A. (2020). *DDSP: Differentiable Digital Signal Processing*. ArXiv. https://doi.org/10.48550/arxiv.2001.04643

Fang, A., Liu, A., Seetharaman, P., & Pardo, B. (2020). *Bach or Mock? A Grading Function for Chorales in the Style of J. S. Bach*. ArXiv. *ArXiv:2006.13329 [Cs.SD]*

Fernandez, J. D., & Vico, F. (2013). AI Methods in Algorithmic Composition: A Comprehensive Survey. *Journal of Artificial Intelligence Research*, 48, pp. 513–582. https://doi.org/10.1613/jair.3908

Fiebrink, R., & Caramiaux, B. (2018). The Machine Learning Algorithm as Creative Musical Tool. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 181–208). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.23

Fiebrink R. & Cook P. R. (2010). The Wekinator: a system for real-time, interactive machine learning in music. *Proceedings of The Eleventh International Society for Music Information Retrieval Conference.*

Fraser, J. (2019). *The voice that calls the hand to write: exploring the adventure of agency and authorship within collaborative partnerships.* Retrieved August 2020, from https://www.julietfraser.co.uk/papers/

Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix Recommender System. *ACM Transactions on Management Information Systems (TMIS)*, *6*(4). https://doi.org/10.1145/2843948

Goodyer, J. (2021, October 14). *How an artificial intelligence finished Beethoven's last symphony.* BBC Science Focus Magazine. https://www.sciencefocus.com/news/ai-beethovens-symphony/

Gotham, M. (2014). Coherence in Concert Programming: A View from the U.K. *IRASM*, *45*(2), pp. 293–309

Hadjeres, G., Pachet, F., & Nielsen, F. (2017). DeepBach: a Steerable Model for Bach Chorales Generation. *Proceedings of the 34th International Conference on Machine Learning*.

Hallström, E., Mossmyr, S., Sturm, B., Vegeborn, V., Wedin, J. 2019. 'From Jigs and Reels to Schottisar och Polskor: Generating Scandinavian-like Folk Music with Deep Recurrent Networks'. In *Sound and Music Computing 2019.*

Han Y., Kim J., and Lee K. 2017. 'Deep convolutional neural networks for predominant instrument recognition in polyphonic music'. *IEEE/ACM Transactions on Audio, Speech and Language processing (TASLP), 25(1),* pp. 208–221

Hao, K. (2019, June 6). *Training a single AI model can emit as much carbon as five cars in their lifetimes.* MIT Technology Review. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Hao, K. (2021, March 11). *He got Facebook hooked on AI. Now he can't fix its misinformation addiction*. MIT Technology Review. https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/

Haworth, C. (2018). Technology, Creativity, and the Social in Algorithmic Music. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 557–582). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.13

Hern, A. (2019, February 14). *New AI fake text generator may be too dangerous to release, say creators.* The Guardian. https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., Hoffman, M. D., Dinculescu, M., & Eck, D. (2018). *Music Transformer.* ArXiv. *arXiv:1809.04281 [cs.LG]*

Hubbard, T. L. (2022). The Pythagorean comma and preference for a stretched octave. *Psychology of Music*, *50*(2), pp. 670–683. https://doi.org/10.1177/03057356211008959

Kanga, Z. (2014). *Inside the Collaborative Process: Realising New Works for Solo Piano* (Doctoral dissertation, Royal Academy of Music, UK)

Ma, B. (2021) *On dialogues between sound and performance physicality: Compositional Experimentation, Embodiment, and Placement of the Self*. (Doctoral dissertation, The Royal Northern College of Music in collaboration with Manchester Metropolitan University, UK)

Miyazaki, S. (2013). AlgoRHYTHMS Everywhere - a heuristic approach to everyday technologies. In J. H. Hoogstad & B. Stougaard Pedersen (Eds.), *Off Beat: Pluralizing Rhythm* (pp. 135–148). Rodopi.

Laidlow, R., & Morris, J. (2020, May 15). *Inventing the Lovelace Engine*. RNCM PRiSM. https://www.rncm.ac.uk/research/research-centres-rncm/prism/prism-blog/inventing-the-lovelace-engine/

Laksnoria, K. (2018). Grammarly as a Tool to Improve Students' Writing Quality: Free Online-Proofreader across the Boundaries. *JSSH (Jurnal Sains Sosial Dan Humaniora)*, *2*(1), pp. 83–89. https://doi.org/10.30595/JSSH.V2I1.2297

Landy, L. (2011). Sound-Based Music 4 All. In *The Oxford Handbook of Computer Music*. Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780199792030.013.0025

Langkjær-Bain, R. (2018). Five ways data is transforming music. *Significance*, *15*(1), pp. 20–23. https://doi.org/10.1111/J.1740-9713.2018.01106.X

Lau, J. (2020). *Google Maps 101: How AI helps predict traffic and determine routes*. Retrieved March 29th 2022 from https://blog.google/products/maps/google-maps-101-how-ai-helps-predict-traffic-and-determine-routes/

Leslie, D. (2020). Understanding Bias in Facial Recognition Technologies. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3705658

Lewis, G. E. "'Is Our Machines Learning Yet?' Machine Learning's Challenge to Improvisation and the Aesthetic," in *Machinic Assemblages of Desire: Deleuze and Artistic Research 3*, edited by Paulo de Assis and Paolo Giudici, pp. 115-128 (Leuven: Leuven University Press, 2021), https://muse.jhu.edu/book/82127.

Lewis, G. E. (2018). Why Do We Want Our Computers to Improvise? In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 123–130). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.29

Li, Y. (2018). *Deep Reinforcement Learning: An Overview*. ArXiv. *arXiv:1701.07274 [cs.LG]*

Lee, L.-H., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., Kumar, A., Bermejo, C., & Hui, P. (2021). *All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda*. ArXiv. https://doi.org/10.48550/arxiv.2110.05352

Liu, A., Fang, A., Hadjeres, G., Seetharaman, P., & Pardo, B. (2020). *Incorporating Music Knowledge in Continual Dataset Augmentation for Music Generation*. ArXiv. *arXiv:2006.13331 [cs.SD]*

Lucidrains (2021). 'BigSleep'. GitHub Code. https://github.com/lucidrains/big-sleep

Lovelace A. A. (Trans.) Sketch of the analytical engine invented by Charles Babbage, with notes by the translator. *Scientific Memoirs, Vol. 3, 1843*, pp. 666-731

Magnusson, T., & McLean, A. (2018). Performing with Patterns of Time. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 245–266). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.21

Matthews, K. (2018). Beyond Me. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 503–506). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.34

McLean, A., & Dean, R. T. (2018a). Musical Algorithms as Tools, Languages, and Partners. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 3–16). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.24

McLean, A., & Dean, R. T. (2018b). Algorithmic Trajectories. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 647–658). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.37

McLeavey, C. (2018). *Clara: A Neural Net Music Generator*. https://christinemcleavey.com/clara-a-neural-net-music-generator/

Mehri S., Kumar K., Gulrajani I., Kumar R., Jain S., Sotelo J., Courville A., Bengio Y. 2017. *SampleRNN: An Unconditional End-To-End Neural Audio Generation Model*. ArXiv. *arXiv:1612.07837v2 [cs.SD]*

Melen, C. (2020). *PRiSM SampleRNN*. RNCM PRiSM. https://www.rncm.ac.uk/research/research-centres-rncm/prism/prism-collaborations/prism-samplernn/

Miller, C. (2018). *Transformative Mimicry: Composition as Embodied Practice in Recent Works* (Doctoral dissertation, University of Huddersfield, UK)

Mogaji, E., Olaleye, S., & Ukpabi, D. (2020). Using AI to Personalise Emotionally Appealing Advertisement. In N. P. Rana, E. L. Slade, G. P. Sahu, H. Kizgin, N. Singh, B. Dey, A. Gutierrez, & Y. K. Dwivedi (Eds.), *Digital and Social Media Marketing* (pp. 137–150). Springer, Cham. https://doi.org/10.1007/978-3-030-24374-6_10

Morton, T. (2013). *Hyperobjects: Philosophy and Ecology after the End of the World*. University of Minnesota Press.

Nierhaus, G. (2009). *Algorithmic composition: Paradigms of automated music generation*. Springer. https://doi.org/10.1007/978-3-211-75540-2

Norrington, R. (2004). The sound orchestras make. *Early Music*, *32*(1), pp. 2–6. https://doi.org/10.1093/EARLYJ/32.1.2

Oore S., Simon I., Dieleman S., Eck D., Simonyan K. 2018. *This Time With Feeling: Learning Expressive Musical Performance*. ArXiv. *arXiv:1808.03715 [cs.SD]*

Pachet, F. (2003). The Continuator: Musical Interaction with Style. *International Journal of Phytoremediation*, *21*(1), pp. 333–341. https://doi.org/10.1076/JNMR.32.3.333.16861

Pachet, F. (2008). The future of content is in ourselves. *Computers in Entertainment*, *6*(3). https://doi.org/10.1145/1394021.1394024

Payne, C. (2019). "MuseNet." *OpenAI*.

Piper, K. (2019, May 15). *A poetry-writing AI has just been unveiled and it's pretty good*. Vox. https://www.vox.com/2019/5/15/18623134/openai-language-ai-gpt2-poetry-try-it

Povilionienė Rima. (2017). Definition Problem of Algorithmic Music Composition. Re-evaluation of the Concepts and Technological Approach. *GESJ: Musicology and Cultural Science*, *16*(2).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). "Language Models are Unsupervised Multitask Learners". *OpenAI*.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. ArXiv. *arXiv:2103.00020 [cs.CV]*

Rohrhuber, J. (2018). Algorithmic Music and the Philosophy of Time. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 17–40). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.1

Rutherford-Johnson, T. (2017). *Music after the fall: Modern composition and culture since 1989*. University of California Press. Pg 62.

Sargeant, A. (2017). The Undeath of Cinema. *The New Atlantis*.

Schreiner, M. (2022, June 12). *DALL-E 2 Test: Can you tell the difference between AI and human art?* Mixed News. https://mixed-news.com/en/dall-e-2-test-can-you-tell-the-difference-between-ai-and-human-art/

SethBling. (2015, June 13). *MarI/O - Machine Learning for Video Games* [video]. YouTube. https://youtu.be/qv6UVOQ0F44

Simoni, M. (2018). The Audience Reception of Algorithmic Music. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 531–556). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.14

Sparkes, M. (2022, June 13). *Sentient AI: Has Google's LaMDA artificial intelligence really come to life?* New Scientist. https://www.newscientist.com/article/2323905-has-googles-lamda-artificial-intelligence-really-achieved-sentience/

Spiegel, L. (1981, January). Manipulations of Musical Patterns. *Proceedings of the Symposium on Small Computers and the Arts*. https://www.researchgate.net/publication/266316606_Manipulations_of_Musical_Patterns

Spiegel, L. (2018). Thoughts on Composing with Algorithms. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 105–112). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.26

Tatalovic, M. (2018). AI writing bots are about to revolutionise science journalism: we must shape how this is done. *Journal of Science Communication*, *17*(1). https://doi.org/10.22323/2.17010501

Torrence, Jennifer (2018). Rethinking the Performer: Towards a Devising Performance Practice. *VIS – Nordic Journal for Artistic Research*

Van den Oord A., Dielemann S., Zen H., Simonyan K., Vinyals O., Graves A., Kalchbrenner N., Senior A., Kavukcuoglu K. 2016. *WaveNet: A Generative Model for Raw Audio*. ArXiv. *arXiv:1609.03499v2 [cs.SD]*

Vickers, B., & Allado-McDowell, K. (2021). *Atlas of Anomalous AI* (B. Vickers & K. Allado-McDowell, Eds.). Ignota Books.

Walshe, J. (2016). *The new discipline: A compositional manifesto*. Retrieved February 2021, from https://musiktexte.de/epages/dc91cfee-4fdc-41fe-82da-0c2b88528c1e.sf/en_US/?ObjectPath=/Shops/dc91cfee-4fdc-41fe-82da-0c2b88528c1e/Products/MT-149

Wang, P., Angarita, R., & Renna, I. (2018). Is this the Era of Misinformation yet: Combining Social Bots and Fake News to Deceive the Masses. *The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018*, pp. 1557–1561. https://doi.org/10.1145/3184558.3191610

Wardhaugh, B. (2017). *Music, Experiment and Mathematics in England, 1653-1705*. Routledge. https://doi.org/10.4324/9781315090764

Wiggins, G., & Forth, J. (2018). Computational Creativity and Live Algorithms. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 267–292). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190226992.013.19

Whittall, A. (2017). Composer-performer collaborations in the long twentieth century. In E. Clarke, & M. Doffman (Eds.) *Distributed Creativity: Collaboration and Improvisation in Contemporary Music.* Oxford University Press

Whorley, R. P., & Laney, R. (2021). Generating Subjects for Pieces in the Style of Bach's Two-Part Inventions. *Proceedings of the 2020 Joint Conference on AI Music Creativity*. https://doi.org/10.30746/978-91-519-5560-5

Musical Works Referenced

Adams, John. *Grand Pianola Music.* 1981.

Adams, John. *Harmonium*. 1981.

Adès, Thomas. *Asyla*. 1997.

Adès, Thomas. *Concentric Paths*. 2005.

Adès, Thomas. *Darknesse Visible*. 1992.

Adès, Thomas. *Piano Quintet*. 2001.

Autreche. *Clipper*. 1995.

Berg, Alban. *Violin Concerto*. 1935.

Birtwistle, Harrison. *Carmen Arcadiae Mechanicae Perpetuum*. 1978.

Birtwistle, Harrison. *Deep Time.* 2016.

Finnis, Edmund. *The Air, Turning.* 2016.

Finnis, Edmund. *Shades Lengthen*. 2015.

Gordon, Michael. *Rewriting Beethoven's Seventh Symphony.* 2006.

Hauff, Helena. *C45p*

Herbert. *Thinking of You*. 1996.

Herndon, Holly. *PROTO*. 2019

Howard, Emily. *Afference*. 2014.

Howard, Emily. *Torus*. 2016.

Knussen, Oliver. *O Hototogisu!*. 2017.

Iannotta, Clara. *Moult*. 2019.

Leith, Oliver. *Honey Siren*. 2020.

Lewis, George. *Minds in Flux.* 2021.

Ligeti, György. *Cello concerto.* 1966.

Luther Adams, John. *Become Desert.* 2018.

Luther Adams, John. *Become Ocean.* 2014.

Luther Adams, John. *Become River.* 2010.

Macklay, Sky. *Many, Many Cadences.* 2016.

Messaien, Olivier. *Quartet for the end of time.* 1941.

Miller, Cassandra. *Guide.* 2013.

Norman, Andrew. *Play.* 2013.

Norman, Andrew. *Sustain.* 2018.

Norman, Andrew. *Switch*. 2015.

Reich, Steve. *Mallet Quartet.* 2010.

Sawer, David. *The Greatest Happiness Principle*. 1997.

Saunders, Rebecca. *Skin.* 2016.

Saunders, Rebecca. *To an utterance*. 2020.

Sciarrino, Salvatore. *la bocca, i piedo, il suono.* 2001.

Sciarrino, Salvatore. *Efebo con radio*. 1981.

Shaw, Caroline. *Plan & Elevation*. 2019.

Snarky Puppy. *Quartermaster*. 2012.

Stravinsky, Igor. *Symphonies of Wind Instruments.* 1920.

Steen-Andersen, Simon. *TRIO.* 2019.

Thorvaldsdottir, Anna. *Aeriality*. 2011.

Walshe, Jennifer. *Ultrachunk.* 2018.

Appendix 1: 'Folk-RNN' and *Three AI Folk Songs*

Three AI Folk Songs for solo violin was commissioned by the Alan Turing Institute for a showcase of UK technical innovation at the British Embassy in Beijing. While the Embassy was interested in showing the confluence between folk tradition and advanced technology, I personally was interested in a test-case of the algorithm FolkRNN. I hadn't used FolkRNN before but was interested in its outputs.

'Folk-RNN' is a symbolic-generative MIDI-based RNN, like 'Clara', specifically trained on transcriptions of folk music. It uses a kind of music-to-text encoding, turning musical data into ABC notation. It produces folk tunes of varying length, usually divided into 2-4 repeating sections of 4, 8, or 16 bars. There is an intuitive web interface that allows users to generate their own tunes, but one creative limitation is the lack of control beyond basic parameters such as tonal area and temperature. I use this in *Three AI Folk Songs* and *Silicon.*

Like *Turing Test // Prelude,* I do not edit the AI-generated material in any way, beyond selecting which three generations to use as a basis for the work and transposing them to a more idiomatic key for the violin.

To me, folk music has a distinct identity not so much through its melodic phrases, harmonic language, and structural fingerprints but through its place in community and communal activities. Since FolkRNN only has scores from which to learn, it necessarily has no understanding of the history or tradition of folk music. I wondered whether an AI can ever truly learn to generate folk music, or whether it will only ever be able to generate music that *sounds like* folk music – and how far that distinction matters to different people. On the other hand, we might consider folk music to be ultimately music which folk musicians perform. In this case, it becomes folk music when performed in that context, even if the material was generated by an AI.

I include this short discussion on *Three AI Folk Songs,* and the score as an Appendix, to provide further context for my inclusion of FolkRNN in *Silicon*, in addition to my development of thought in the area of what constitutes authenticity in music.

Appendix 2: Style Transfer and 'NSynth'

For the Royal Northern College of Music's 'Spotlight' Concert series in December 2019, I organised a short concert experiment on the theme of 'Style Transfer'. Style Transfer is a computer vision technique often used to take two images, a source and a target, and blend them together, such that the source is rendered in the style of the target. Recently, there has also been work on style transfer in the audio domain. This usually involves taking using one sound's spectral profile as a target and an audio input as a source. Machine learning is then used to reconstruct the source to sound like the target.

I was interested in this technology and was keen to experiment with it in an informal setting. I wanted to concentrate on quickly getting a feel for audio-domain style transfer in a musical setting without needing to dedicate too much time to notating music or troubleshooting code. For this reason, I chose to use Google Magenta's 'NSynth' algorithm (Engel et al 2017), which can be used in real-time on a regular laptop. 'NSynth' is a synthesizer that uses a 'WaveNet'-style AI to generate individual samples, building this into sound. Theoretically, it allows easy and quick creation of synthesized timbres that would be difficult or impossible to generate through traditional methods. The interface I worked with allowed me to use machine learning to combine the waveform profiles of different instruments and create new waveforms.

I also decided this performance should be partly improvised, with only basic musical materials written out for myself (performing on 'NSynth') and two string players. I am a trained jazz improviser but had not engaged in this area of my practice for several years. Improvising with 'NSynth' showed me both the creative possibilities of this technology, which I immediately resolved to include in a larger-scale work in the future, and the advantages of having an interface that facilitates the skills of the individual.

In terms of interface, my improvisation skills are on the saxophone, and it was necessary to interact with ‘NSynth’ through a keyboard synthesizer which I am less familiar with in this context. For future projects, I wanted to ensure the interface between performer and style transfer algorithm was as natural as possible, especially if I were to be using it with orchestral musicians who have very limited rehearsal time. As an example, for me this might have involved finding some way to incorporate the saxophone into the interface either directly as live audio or by transforming this audio stream into another form of data.

Appendix 3: ‘GPT-2’ and ‘WaveNet’ texts during *Alter*

Selected from the following generations by ‘GPT-2’. Words in **bold** were the prompt given to ‘GPT-2’. Selections were then recited by ‘WaveNet’ as part of the piece’s electronics track.

The engine might compose a very interesting engine in its own right, but we also have a lot of interest in having an engine capable of working on the world as we are experiencing it," he says. "[It also means] that you'll be able to do some fantastic things with it."

With this in mind, the team turned its attention to the problem posed by "unrealistic" gameplay in Star Wars: Battlefront II. "As with most things, we're trying to find the best solution," says Puharich. If you remember from that brief preview about the "cave man", "we're interested in making a little less dramatic space battles in the game," because players "have more reason than ever to feel as though they're playing on a world we've built," he adds.

The engine might compose itself, even if it doesn't. In the real world, we don't spend much time looking for the wrong part or the right kind of part. We just use parts we know we're gonna need.

If we want to make a good engine, we actually need the right parts. It goes beyond just parts. If the same part is being used on two different parts, we might need to do something like change their internals or tweak their alignment: if they're not performing perfectly. A good engine is designed to ensure that no engine component ever fails. This is called good faith. If someone starts by replacing a bad, badly aligned component, we're still good to go. But if the component is actually performing flawlessly after modification, then it isn't a problem—we do as we ought.

Here's what makes parts useful: the things the engine doesn't care about, like the ignition-cannon motor and gearbox, aren't really engine components. These are also often just like the parts to get you in and out of the car and make sure you can drive.

Now, if the engine was designed based on the needs of the real world—but it always had to compete with people who already had an engine that could work as a whole with a low-quality car and with parts that were a little less than perfect

The engine might compose a single unit of high-performance power but it has no way to perform more than one work at once. If you were to build the engine with only three units, they would be quite a large machine, but only about 20 of them were needed for the prototype project. To add the capacity of the machine at this scale, you would have to build the engines from two different steel frames. And that is too much complexity for an engine which is intended primarily for small and fast cars, and which was never intended for any kind of huge scale. So we decided not to create this engine with the intention to scale up and take on the world, but rather to build it for the maximum cost.

The engine might compose itself as a single unit without the need for external components to function properly; this is referred to as a non-integral unit.

The diagram shows the output (see above). The red line shows the power and noise levels. The horizontal line has an angle in it at the point of clipping. In the graph above we show a straight line. Note how the noise is different as a function of the angle in the right line. As we move along the right line, the noise grows. At 90° angle the noise remains static as noise level. At 180° the noise rises but then it gradually drops as angle increases. The graph below shows both power and noise at this angle.

The engine might compose a single unit unit, while that unit unit could consist of additional components, thus causing them to act in multiple ways. For example, a piston might act as an internal combustion engine, while other internal combustion engines would be integrated inside it. The internal combustion engine will need to be cooled and powered by a fan. The exhaust exhaust gases (or gases within a cylinder) should evaporate and exit the cylinders to allow for further exhaust velocity.

The internal combustion engine will need to be cooled and powered by a fan. The exhaust exhaust gases (or gases within a cylinder) should evaporate and exit the cylinders to allow for further exhaust velocity. A propeller will turn like a propeller. With a propeller, a force (tricycle force) on its side will apply a force on its forward end which produces thrust (propeller pull).

An oscillating engine generator, for example a propeller generator, produces thrust just by applying additional thrust.

The engine might compose well of three elements:

The engine might compose music. He may use it to play games with our friend's wife to keep her company… he may compose music for our game or movie. He might compose an entire novel or play on stage.

We can only imagine what else.

The engine might compose music, and it might do a lot of crazy sounds and things like that, but when it moves, it'll be more like this.

The engine might compose music, but it may also create new ones from them, which it can do in two ways: it may generate these new pieces of music independently of the original material created by it, or it may make them from scratch. This is a process referred to as compositional composality. In the case of compositional composality, the composition of the pieces of music depends on the number and quality of the pieces of music used. The quality of the music depends on what the composition is meant to be.

The engine might compose music in a different way, but I haven't tried it yet—I'll have to test that. I don't know if I'll ever use it, because once I start using it, my mind goes completely backwards, and I will probably never do anything again. But if it does, I'll be very happy. I won't regret that, at all, because of the quality of the sound that I created.

The engine might compose music like your favourite electronic music, but I think I can do better when I think about it, because I don't really look at it as an engine. It's rather about being open-ended and flexible

The engine might compose elaborate and scientific pieces of music of any degree of complexity or extent. There's no doubt that, in some cases, it could inspire music by itself. You could write a poem about the engine, or a symphony that had an element of the engine. One might also write a song or ballet that had an element of the engine. That's all true. There's no denying that it's possible: You could compose a song or musical piece that you could express through the power of the engine. To me, that's just too much: It's too much, really. I don't think that it'd be right, of course, for a composer to compose a poem without the power of the engine, or to write a symphony without the power of the engine, or to write an opera without the power of the engine. In fact, it would be silly, for two reasons. First, if you can compose poetry or music, you can write music, and that's true with any genre that's popular right now. You don't need the engine to do it. But people should write something that's entertaining without the engine. And of course, they don't. You wouldn't like to hear a movie without the engine. If you don't enjoy movies, you can't like to watch them. In any case, you've got to appreciate the pleasure of music

The engine might compose elaborate and scientific pieces of music of any degree of complexity or extent. It would have to go very, very very far above what we see in the movies and movies that use machines, which would be some of the most advanced machines imaginable. I think you'd be surprised how far I can imagine the engine of an airplane getting."

The engine might compose elaborate and scientific pieces of music of any degree of complexity or extent. But like the other instruments on a symphonic orchestra, it is capable of a remarkable amount of expression and musicality—a rare feature of music. It has a remarkable ability, for instance, to develop and elaborate, in some very subtle ways, a series of complex forms without any obvious way for a mere conductor or orchestraman to know them.

The engine might compose elaborate and scientific pieces of music of any degree of complexity or extent. It might be a combination of musical instruments; like the piano- flute, violin or harpsichord. But it will still be a musical instrument and not just a musical instrument, and any human being would feel the need to master these instruments. But how could they play any great degree of complex music which was beyond the limits of human abilities? Would they be able to create melodies that were of such an extent of sophistication that they could be played at the level of the human brain?

If you thought music was something like the music of the orchestra, that was an entirely different question. Music is composed of sounds and sounds are composed of things and things are composed of other things.

The engine might compose elaborate and scientific pieces of music of any degree of complexity or extent. So how could there be something so difficult, so esoteric, so exotic, so difficult, so impossible to understand? I was sure it was because the artist himself was unable to comprehend how to make something that was entirely rational - it seemed he didn't have the ability!

The artificial intelligence might compose elaborate and scientific pieces of music of any degree of complexity or extent. That is to say, any type of intelligence might compose a composite piece or musical composition based on different parts. As well, if a robot is built like a computer, it would not have any notion of "right" or "wrong" when it comes down to composing complex pieces of music.

The artificial intelligence might compose elaborate and scientific pieces of music of any degree of complexity or extent. But it's still working with a melody or sound, like a cello, or guitar, or anything else.

The artificial intelligence might compose elaborate and scientific pieces of music of any degree of complexity or extent. It can produce even greater patterns than human music.

The artificial intelligence might compose elaborate and scientific pieces of music of any degree of complexity or extent. It might understand how to put together beautiful compositions of any range of musical complexity of any length and range of pitch or volume. It might interpret what we mean and to what extent the words and thoughts that we say are a part of some sort of universal language. It might also have what the brain has to do: to recognize an image and recognize it from its particular context.

The artificial intelligence might compose elaborate and scientific pieces of music of any degree of complexity or extent. Its creativity can be quite extraordinary – some seem to have gone beyond human nature, and have a mind so complex that some even perceive that they can hear the music of their subconscious thoughts as if they were themselves musical.

Appendix 4: Musical Time Array in *Gravity*

*The **frequency** of **discernible** events that **traverses** an **array** in one direction*

For me, it is the combination of these four highlighted words that dictates the *sense* of time flowing in a piece of music, the feeling that the music is closer to or further from stasis or spiralling acceleration. In this context, discernible means "not audibly identical", to traverse means to "move audibly through (i.e. changing) one of the other three arrays (tuning/timbre/texture). Thus, according to this definition, music should feel in some way timeless or in stasis if audible events are indiscernible from one another, are not changing to a new position on any array and are infrequent.

By contrast, music should feel fast-flowing if there is a very high frequency of events that are all audibly distinct from one another, that are moving to new positions in one or more arrays. All three must be true: a high frequency of discernible events is not enough to constitute time flowing if no array is traversed, or if the music is moving forwards and backwards through the same positions in an array. I think this point was grasped by common practice era composers, substituting "array" for "harmonic spectrum". To me, music that frequently changes between I and V (two discernible harmonic events) does not feel like it is traversing time in the same way that music that is very quickly modulating does. According to this definition, perception of musical time has no direct relation to metre, rhythm or tempo.

If this definition of musical time is taken at face value, at least for *Gravity*, it mirrors general relativity in one more way. The other three arrays are, for the most part, independent of one another. An extreme timbre does not necessitate an extreme tuning. However, music that is at the extreme of the time array, say at the extremely "slow end", requires the other arrays to not be traversed. They can be directly bound in this way to the time array. Or, put another way, musical time could be understood to emerge organically from the behaviour of the other three arrays, depending on how fast they are traversing. This is comparable to the three spatial dimensions and one time dimension in general relativity, where extreme curvature of the three affects the last.