Addressing the Gap Between Machine Learning and Casual Music Creation

CS39440 Major Project Report

Author: Daniel Swift (das82@aber.ac.uk)

Supervisor: Dr Chuan Lu (cul@aber.ac.uk)

01st May 2022

Version 1.4 (Draft)

This report is submitted as partial fulfilment of a BSc degree in  
Computer Science (G401)

Department of Computer Science

Aberystwyth University

Aberystwyth

Ceredigion

SY23 3DB

Wales, UK

Declaration of originality

I confirm that:

* This submission is my own work, except where clearly indicated.
* I understand that there are severe penalties for Unacceptable Academic Practice, which can lead to loss of marks or even the withholding of a degree.
* I have read the regulations on Unacceptable Academic Practice from the University’s Academic Registry (AR) and the relevant sections of the current Student Handbook of the Department of Computer Science.
* In submitting this work, I understand and agree to abide by the University’s regulations governing these issues.

Name Daniel Swift

Date 28/04/2022

Consent to share this work

By including my name below, I hereby agree to this project's report and technical work being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Name Daniel Swift

Date 28/04/2022

Acknowledgements

I am grateful to…

I’d like to thank…

Abstract

Music is one of the most expressive and emotive forms of human art. Deeply ingrained in the roots of different cultures and communities, music is widely created and enjoyed in various forms throughout the world. The rise of personal computers and innovations such as the Musical Instrument Digital Interface (MIDI) standard has born a new era of digital music creation, capitalised on from the biggest artists in the world perfecting their sound, to curious users experimenting with music scales in their bedroom.

Coinciding with this is the rise in machine learning, which is also becoming more accessible to a wider, non-specialised audience. However, as of this report, there are still many areas of machine learning that are too complicated or expensive for the more casual user. My project has tried to address this problem, allowing people who may have some background in digital music creation but none with machine learning or AI in general, and allow them to combine these relatively new fields as seamlessly as possible.

In this project I will compare different generative algorithms regarding their suitability for music generation. Furthermore, I will attempt to create software that is user-friendly and intuitive whilst enabling users to generate music from different complex music generation algorithms.

Contents

1. Background, Analysis & Process 7

1.1. Background 7

1.1.1. Initial Motivations 7

1.1.2. Music and Sound Research 7

1.1.3. MIDI Standard and Useful Filetypes 8

1.1.4. Neural Network Models 8

1.1.5. Similar Existing Solutions 9

1.2. Relevant Tools and Libraries 10

1.2.1. Keras/TensorFlow 10

1.2.2. PyTorch 10

1.2.3. Scikit-Learn 11

1.2.4. Magenta 11

1.2.5. Anaconda 11

1.3. Analysis 11

1.4. Process 12

2. Machine Learning for Music Generation 13

2.1. Introduction 13

3. Design 13

3.1. Overall Architecture 13

3.2. Detailed Design 13

3.2.1. Even More Detail 13

3.3. User Interface Design 14

3.4. Other Relevant Sections 14

4. Implementation 15

5. Testing 16

5.1. Overall Approach to Testing 16

5.2. Automated Testing 16

5.2.1. Unit Tests 16

5.2.2. User Interface Testing 16

5.2.3. Stress Testing 16

5.2.4. Other Types of Testing 17

5.3. Integration Testing 17

5.4. User Testing 17

6. Critical Evaluation 18

7. References 19

8. Appendices 21

A. Third-Party Code and Libraries 22

B. Code Samples 23

# Background, Analysis & Process

This section should discuss your preparation for the project, including background reading, your analysis of the problem and the process or method you have followed to help structure your work. It is likely that you will reuse part of your outline project specification, but as you write this report at the end of the project you should have more to discuss.

**Notes**:

* All of the sections and text in this example are for illustration purposes. The main Chapters are a good starting point, but the content and actual sections that you include are likely to be different.
* Look at the document MMP\_SO8 Project Report and Technical Work **Error! Reference source not found.** for additional guidance.

## Background

### Initial Motivations

My initial interest in pursuing this area for my project stemmed firstly from the fact that I love both listening and playing guitar music. I started playing guitar around five years ago, and have accrued some knowledge in how to write my own music and some of the theory behind it. Furthermore, I believe that in those five years, the way that I perceive and understand music that I listen to has changed drastically. Being able to play a musical instrument has made me appreciate skill and creativity a lot more, especially guitar music.

This is no coincidence, as musicians with different expertise selectively engage different sensorimotor regions in the brain when listening to the instruments they are able to play [1]. However, this is not to say that people with some kind of musical expertise are better at perceiving music than non-musicians, obviously non-musicians are able engage, understand, and build powerful connections with music they listen to. This is shaped by numerous factors, such as mood, setting, the piece itself[2], or the person’s familiarity with the piece and genre[3].

In terms of familiarity with machine learning, I spent my year in industry working in the marketing department for a company producing machine learning and AI related products, with emphasis on these products being relatively easy to program and deploy within businesses and schools. The focus on making AI and machine learning more user-accessible and widely accepted was another source of inspiration for me to partake in this specific project. I hoped that it might be an insightful way to get more people into both writing music and machine learning, as well as being a genuinely useful tool for basic music composition.

### Music and Sound Research

My initial research and reading predominantly regarded the different kind of neural network models that could be used for this project, the libraries that supported them, the languages or programming interfaces that I could use, and similar existing solutions that I could draw inspiration from.

Music itself is composed of many different elements, which combine to create the various melodies that we can perceive in a variety of ways. It can be conveniently described as a series of notes, which play out in a linear pattern from the beginning of a song until the end. These notes can be identified by their pitch (the frequency of the soundwave they produce), duration, velocity, volume, and many others. Different notes can be played at the same time to make up chords, and can be played in conjunction with other specific notes to make up scales.

It is important to understand what makes music sound “good”. Sure, pressing random keys on a piano or plucking at random strings on a guitar is probably not going to sound like a masterpiece, but in order to play something harmonious, notes need to stay in the same key, and make sure they are played within defined scale intervals.

Leading on, it could be easy enough to develop an algorithm where different scales are hardcoded, and it would select notes at random (or in a loosely determined pattern based on other inputs) in order to generate a piece that, while potentially sounding quite weird, would be in key. This would, in turn, reduce the sense of originality, as the piece would be the product of either over-determinism or pure randomness, where in reality we would require more of a balance.

This would justify a need for machine learning, as, for instance, a neural network model could be trained from music that is generally well regarded, and would likely follow different scale intervals and remain in key. The model would pick up on this so that when it generates its own music, the notes should all remain in key. However, remaining in key is not always the most important factor. It is worth acknowledging that notes can sound good sometimes when played out of key, and this should be reflected in the songs within the provided training datasets. Furthermore, obviously the concept of “good” music is subjective, and different individuals will form an impression based on the reasons aforementioned in 1.1.1.

### MIDI Standard and Useful Filetypes

MIDI is a technological standard that evolved to enable communication between digitally enabled instruments, such as 1980’s synthesisers. Later, it provided a platform for users with personal computers to record, store and edit music[9]. MIDI files themselves are normally relatively small, meaning that they do not require much disk space, due to the fact that they do not have any actual audio data. In place of this, they contain data regarding what notes are to be played, the time that they are played at, the duration of the note, and how loud it should be[10]. Due to this, many musical pieces that are converted to MIDI files will sound different to the original recording, and sometimes unpleasant.

### Neural Network Models

In terms of selecting a specific type of generation algorithm, there were a few options. Most early music generation techniques used Recurrent Neural Networks (RNN), such as Long Short-Term Memory (LSTM) networks, which would generate single-instrument music in the same way that natural language models work[4][5][6].

A big benefit of using RNNs is that it has the ability to apply contextual information when mapping between input and output sequences, however the drawback is that the range of this context is limited. To explain this further, within RNN architectures, the influence of a given input on the hidden layer, and then in consequence the output, either decays or blows up exponentially as it cycles through the network’s recurrent connections[7]. This is more commonly known as the vanishing gradient problem[8].

LSTMs have been developed to address these limitations, and have proven more successful at it[12]. Within the architecture of an LSTM is a constant error carrousel (CEC), which basically affords protection to input gates by perturbing irrelevant inputs [13].

Another potential model could be a Generative Adversarial Network (GAN). GANs work using unsupervised learning, however it frames itself as a supervised learning algorithm, using two sub-models, a generator model, and a discriminator model. The generator model is trained to generate new data from an existing dataset, and the discriminator model attempts to classify data that is real (from the given dataset), or generated by the generator model. The two models work at the same time until the discriminator fails to correctly determine the difference between real and generated data at a pre-determined rate[19], the original proposed rate being 50%[20].

### Similar Existing Solutions

#### SOUNDRAW

A screenshot of a computer

Description automatically generated with medium confidence“Soundraw” is a web-based music generator that uses pre-trained AI to generate songs at a selection of lengths, based on a user-given input of the song’s mood, genre(or style), tempo and instrument. The generated music can also be edited to cut its length, slow or quicken its beats per minute (bpm), change the key, and change the volume, amongst other advanced attributes at specific timings.

Figure 1: The GUI for soundraw.io[22], with its extensive plethora of features

The design of this app is very professional and fairly straightforward, and its use of well-developed pre-trained models is a good source of inspiration for how this project could work. However, this solution offers no way for users to train their own models, or select the different kind of generation algorithm they would like to use.

#### dopeloop.ai

The “dopeloop” melody generator is a web-based application that is able to generate melodies from three user-given inputs. The first is how many notes to generate (the length), the second the root note, and the third is the scale to generate the notes in. The notes are generated randomly within these given confinements. The user also has the option to change the instrument and the bpm. This application is also available as a mobile app.

Graphical user interface, application

Description automatically generated

Figure 2: The GUI for dopeloop.ai's[21] melody generator

Whilst the way that music is generated for this solution is not the same way that it will be generated in this project, the simplicity of the GUI design and the swiftness that it operates provides a large source of inspiration for the practical outcome of this project, as well as advocating for a web-based solution.

## Relevant Tools and Libraries

### Keras/TensorFlow

TensorFlow is an open-source library built specifically to allow developers to create and train machine learning models. Created by Google, TensorFlow is predominantly used with Python and JavaScript, and it is one of the most widely used libraries in relation to machine learning[11]. Keras is an extension of TensorFlow which tries to make using it a little more user friendly.

### PyTorch

Similarly to TensorFlow, PyTorch is an open-source library, but focused on accelerated deep neural network programming and developed by Facebook[14]. However, when compared to TensorFlow, it would probably fall a little short for this project. This is because TensorFlow has a superior and well-documented framework, more of an abundance of tutorials, and better visualisation of its processes, whereas PyTorch provides only limited visualisation. On the other hand, PyTorch gains optimal performance with native asynchronous execution within Python, whereas for TensorFlow this would have to be manually coded in[15].

### Scikit-Learn

Scikit-Learn offers a wide range of machine learning algorithms, with alternatives to neural networks. Decision trees, Support Vector Machines (SVM), and linear regression, to name a few, are all available to use through this library.

### Magenta

Magenta is a self-described open-source research project, created by Google using the TensorFlow library. It is specifically designed to explore the role of machine learning when undergoing the process of creating art and music, and is available for both Python and JavaScript[16]. Magenta can be used to train custom models, but also comes with a selection of pretrained models, of which can be changed certain parameters to affect the final output of the model[17].

### Anaconda

Anaconda is an open-source Python distribution platform, with a focus on making data science and machine learning easily accessible on one machine. It provides a user-friendly graphical user interface (GUI) which makes it easy to switch between different development environments, and features it own repository of thousands of open-source data science and machine learning related packages. This includes previously mentioned packages such as TensorFlow, PyTorch and Keras[18].

## Analysis

Taking into account the problem and what you learned from the background work, what was your analysis of the problem? How did your analysis help to decompose the problem into the main tasks that you would undertake? Were there alternative approaches? Why did you choose one approach compared to the alternatives?

There should be a clear statement of the objectives of the work, which you will evaluate at the end of the work.

In most cases, the agreed objectives or requirements will be the result of a compromise between what would ideally have been produced and what was determined to be possible in the time available. A discussion of the process of arriving at the final list is usually appropriate.

As mentioned in the lectures, think about possible security issues for the project topic. Whilst these might not be relevant for all projects, do consider if there are relevant for your project. Where there are relevant security issues, discuss how they will this affect the work that you are doing. Carry forward this discussion into relevant areas for design, implementation and testing.

## Process

You need to describe briefly the life cycle model or research method that you used. You do not need to write about all of the different process models that you are aware of. Focus on the process model that you have used. It is possible that you needed to adapt an existing process model to suit your project; clearly identify what you used and how you adapted it for your needs.

# Machine Learning for Music Generation

## Introduction

## Data

## RNNs

## LSTMs

## GANs

# Design

You should concentrate on the more important aspects of the design. It is essential that an overview is presented before going into detail. As well as describing the design adopted it must also explain what other designs were considered and why they were rejected

The design should describe what you expected to do and might also explain areas that you had to revise after some investigation.

Typically, for an object-oriented design, the discussion will focus on the choice of objects and classes and the allocation of methods to classes. The use made of reusable components should be described and their source referenced. Particularly important decisions concerning data structures usually affect the architecture of a system and so should be described here.

How much material you include on detailed design and implementation will depend very much on the nature of the project. It should not be padded out. Think about the significant aspects of your system. For example, describe the design of the user interface if it is a critical aspect of your system, or provide detail about methods and data structures that are not trivial. Do not spend time on long lists of trivial items and repetitive descriptions. If in doubt about what is appropriate, speak to your supervisor.

You should also identify any support tools that you used. You should discuss your choice of implementation tools - programming language, compilers, database management system, program development environment, etc.

Some example sub-sections may be as follows, but the specific sections are for you to define.

## Overall Architecture

## Detailed Design

### Even More Detail

## User Interface Design

## Other Relevant Sections

# Implementation

The implementation should discuss any issues you encountered as you tried to implement your design. During the work, you might have found that elements of your design were unnecessary or overly complex; perhaps third-party libraries were available that simplified some of the functions that you intended to implement. If things were easier in some areas, then how did you adapt your project to take account of your findings?

It is more likely that things were more complex than you first thought. In particular, were there any problems or difficulties that you found during implementation that you had to address? Did such problems simply delay you or were they more significant?

You can conclude this section by reviewing the end of the implementation stage against the planned requirements.

# Testing

Detailed descriptions of every test case are definitely not what is required in this section; the place for detailed lists of tests cases is in an appendix. In this section, it is more important to show that you adopted a sensible strategy that was, in principle, capable of testing the system adequately even if you did not have the time to test the system fully.

Provide information in the body of your report and the appendix to explain the testing that has been performed. How does this testing address the requirements and design for the project?

How comprehensive is the testing within the constraints of the project? Are you testing the normal working behaviour? Are you testing the exceptional behaviour, e.g. error conditions? Are you testing security issues if they are relevant for your project?

Have you tested your system on “real users”? For example, if your system is supposed to solve a problem for a business, then it would be appropriate to present your approach to involve the users in the testing process and to record the results that you obtained. Depending on the level of detail, it is likely that you would put any detailed results in an appendix.

Whilst testing with “real users” can be useful, don't see it as a way to shortcut detailed testing of your own. Think about issues discussed in the lectures about until testing, integration testing, etc. User testing without sensible testing of your own is not a useful activity.

The following sections indicate some areas you might include. Other sections may be more appropriate to your project.

## Overall Approach to Testing

## Automated Testing

### Unit Tests

### User Interface Testing

### Stress Testing

### Other Types of Testing

## Integration Testing

## User Testing

# Critical Evaluation

Examiners expect to find a section addressing questions such as:

* Were the requirements correctly identified?
* Were the design decisions correct?
* Could a more suitable set of tools have been chosen?
* How well did the software meet the needs of those who were expecting to use it?
* How well were any other project aims achieved?
* If you were starting again, what would you do differently?

Other questions can be addressed as appropriate for a project.

The questions are an indication of issues you should consider. They are not intended as a specification of a list of sections.

The evaluation is regarded as an important part of the project report; it should demonstrate that you are capable not only of carrying out a piece of work but also of thinking critically about how you did it and how you might have done it better. This is seen as an important part of an honours degree.

There will be good things in the work and aspects of the work that could be improved. As you write this section, identify and discuss the parts of the work that went well and also consider ways in which the work could be improved.

In the latter stages of the module, we will discuss the evaluation. That will probably be around week 9, although that differs each year.

# References

This final section should list all relevant resources that you have consulted in researching your project.

1. Krishnan, S., Lima, C.F., Evans, S., Chen, S., Guldner, S., Yeff, H., Manly, T. and Scott, S.K., 2018. Beatboxers and guitarists engage sensorimotor regions selectively when listening to the instruments they can play. *Cerebral Cortex*, *28*(11), pp.4063-4079.
2. Blood, A.J. and Zatorre, R.J., 2001. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the national academy of sciences*, *98*(20), pp.11818-11823.
3. Leaver, A.M., Van Lare, J., Zielinski, B., Halpern, A.R. and Rauschecker, J.P., 2009. Brain activation during anticipation of sound sequences. *Journal of neuroscience*, *29*(8), pp.2477-2485.
4. Tham, I., 2021. *Generating Music using Deep Learning*. [online] Medium. Available at: <https://towardsdatascience.com/generating-music-using-deep-learning-cb5843a9d55e#:~:text=Recently%2C%20Convolutional%20Neural%20Networks%20(CNNs,convolutions%20to%20generate%20raw%20audio.> [Accessed 13 March 2022].
5. Skuli, S., 2017. *How I Built A Lo-fi Hip-Hop Music Generator*. [online] Medium. Available at: <https://ai.plainenglish.io/building-a-lo-fi-hip-hop-generator-e24a005d0144> [Accessed 21 March 2022].
6. Medium. 2020. *How I Built A Lo-fi Hip-Hop Music Generator*. [online] Available at: <https://ai.plainenglish.io/building-a-lo-fi-hip-hop-generator-e24a005d0144> [Accessed 19 March 2022].
7. Graves, A., 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pp.37-45.Hochreiter, S., 1998. Recurrent neural net learning and vanishing gradient. *International Journal Of Uncertainity, Fuzziness and Knowledge-Based Systems*, *6*(2), pp.107-116.
8. Hochreiter, S., 1998. Recurrent neural net learning and vanishing gradient. *International Journal Of Uncertainity, Fuzziness and Knowledge-Based Systems*, *6*(2), pp.107-116.
9. University, I., 2022. *The MIDI Standard: Introduction to MIDI and Computer Music: Center for Electronic and Computer Music: Jacobs School of Music*. [online] Cecm.indiana.edu. Available at: <https://cecm.indiana.edu/361/midi.html> [Accessed 29 March 2022].
10. Fisher, T., 2022. *What's a MIDI File and How Do You Open One?*. [online] Lifewire. Available at: <https://www.lifewire.com/midi-file-2621979> [Accessed 13 April 2022].
11. Yegulalp, S., 2019. *What is TensorFlow? The machine learning library explained*. [online] InfoWorld. Available at: <https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html> [Accessed 17 April 2022].
12. Graves, A. and Schmidhuber, J., 2005, July. Framewise phoneme classification with bidirectional LSTM networks. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.* (Vol. 4, pp. 2047-2052). IEEE.
13. Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.
14. Yegulalp, S., 2017. *Facebook brings GPU-powered machine learning to Python*. [online] InfoWorld. Available at: <https://www.infoworld.com/article/3159120/facebook-brings-gpu-powered-machine-learning-to-python.html> [Accessed 21 April 2022].
15. Terra, J., 2022. [online] Available at: <https://www.simplilearn.com/keras-vs-tensorflow-vs-pytorch-article#:~:text=TensorFlow%20is%20an%20open%2Dsourced,because%20it's%20built%2Din%20Python.> [Accessed 13 April 2022].
16. Magenta. 2022. *Magenta*. [online] Available at: <https://magenta.tensorflow.org/> [Accessed 10 April 2022].
17. Sidor, S., 2020. *magenta/magenta/models at main · magenta/magenta*. [online] GitHub. Available at: <https://github.com/magenta/magenta/tree/main/magenta/models> [Accessed 10 April 2022].
18. Dominodatalab.com. 2022. *What is Anaconda? | Domino Data Science Dictionary*. [online] Available at: <https://www.dominodatalab.com/data-science-dictionary/anaconda> [Accessed 25 April 2022].
19. Brownlee, J., 2019. *A Gentle Introduction to Generative Adversarial Networks (GANs)*. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/> [Accessed 24 April 2022].
20. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, *27*.
21. McCormick, C., n.d. *Melody Generator*. [online] Melody Generator. Available at: <https://dopeloop.ai/melody-generator/?s=7019330677856559> [Accessed 1 May 2022].
22. Soundraw.io. n.d. *Soundraw*. [online] Available at: <https://soundraw.io/#about-soundraw> [Accessed 1 May 2022].

# Appendices

The appendices are for additional content that is useful to support the discussion in the report. It is material that is not necessarily needed in the body of the report, but its inclusion in the appendices makes it easy to access.

If you have used any 3rd party code, i.e. code that you have not written yourself such as libraries, then you must include Appendix A. In that appendix, you will provide details of the 3rd party code that you have used.

For most other items, it would be better to include them in your technical submission instead of including them as an appendix. For example:

* If you have developed a Design Specification document as part of a plan-driven approach for the project, then it would be appropriate to include that document in the technical work. In this report, you would highlight the most interesting aspects of the design, referring your reader to the full specification for further detail.
* If you have taken an agile approach to developing the project, then you may be less likely to have developed a full requirements specification at the start of the project. Perhaps you used stories to keep track of the functionality and the ‘future conversations.’ If it isn’t relevant to include all those stories in the body of your report, you could detail those stores in a document in the technical work.
* If you have used manual testing, then include a document in the technical work that records the tests that have been done. In this report, you would talk about the use of those tests.

Documents included in the technical work or in the appendices are supporting evidence of the work done. Where you include documents, this report should refer to the documents. You should not be relying on detailed study of those documents in order to understand what is written in this report.

Speak to your supervisor or the module coordinator if you have questions about this.

* 1. Third-Party Code and Libraries

If you have made use of any third-party code or software libraries, i.e. any code that you have not designed and written yourself, then you must include this appendix.

As has been said in lectures, it is acceptable and likely that you will make use of third-party code and software libraries. If third-party code or libraries are used, your work will build on that to produce notable new work. The key requirement is that we understand what your original work is and what work is based on that of other people.

Therefore, you need to clearly state what you have used and where the original material can be found. Also, if you have made any changes to the original versions, you must explain what you have changed.

The following is an example of what you might say.

**Apache POI library** – The project has been used to read and write Microsoft Excel files (XLS) as part of the interaction with the client’s existing system for processing data. Version 3.10-FINAL was used. The library is open source and it is available from the Apache Software Foundation [5]. The library is released using the Apache License [6]. This library was used without modification.

Include as many declarations as appropriate for your work. The specific wording is less important than the fact that you are declaring the relevant work.

* 1. Code Samples

This is an example appendix. Include as many appendices as you need. The appendices do not count towards the overall word count for the report.

For some projects, it might be relevant to include some code extracts in an appendix. You are not expected to put all of your code here - the correct place for all of your code is in the technical submission that is made in addition to the Project Report. However, if there are some notable aspects of the code that you discuss, including that in an appendix might be useful to make it easier for your readers to access.

As a general guide, if you are discussing short extracts of code then you are advised to include such code in the body of the report. If there is a longer extract that is relevant, then you might include it as shown in the following section.

Only include code in the appendix if that code is discussed and referred to in the body of the report.

Random Number Generator

The Bayes Durham Shuffle ensures that the pseudo random numbers used in the simulation are further shuffled, ensuring minimal correlation between subsequent random outputs.

// Some example code here…