

University of Southampton

Individual Project Plan

Harry Twigg

Submitted in fulfilment of the requirements the Individual Project
at the University of Southampton, October 2021

Contents

1	Proposal	2
1.1	Background	2
1.2	Aim	2
1.3	Objectives	3
1.4	Evaluation Methods	3
2	Timeline	5
3	Other Requirements	6
3.1	Training Needs	6
3.2	Risk Assessment	6
3.3	Resource Access	7

Chapter 1

Proposal

1.1 Background

Machine Learning is rapidly transforming audio processing, yielding wide ranging applications in business and for artists, amongst others. One particularly interesting area is the use of deep based learning methods for music sound synthesis.

Machine Learning methods can already take musical or other sound data input, process it, then provide a variety of potentially useful features, these include:

- Timbre transfer eg. changing a melody from being in one instrument to another
- Pitch transfer eg. transposing an instrument up an octave
- Changing the room accoustics
- Generation of music and speach either randomly or from a given input

1.2 Aim

It is foreseeable that some machine learning sound synthesis methods have potential to be built on further, whilst others may be deemed obsolete. This project aims to compare these methods,

evaluating the advantages and disadvantages of how audio signals are encoded and presented to them. In an optimistic scenario, new methods of audio signal encoding and deep learning methods for sound synthesis shall be researched and proposed.

1.3 Objectives

- Evaluate existing methods of audio signal encoding with a focus on the following machine learning areas and the way data is presented to the network (but not limited to):
 - Frequency based Fourier coefficient methods eg. Tacotron, GANSynth
 - Autoregressive whole single sample methods eg. Wavenet, SampleRNN
 - DDSP a novel modular approach to sound synthesis
- Investigate potentially innovative new methods of sound synthesis and encoding, building on existing researched work, this could include:
 - The use of a new encoding method or formatting of the input signal
 - Building on top of existing sound Synthesis Models, potentially making use of alternative ways of encoding and presenting data to the deep network.
 - Finding a way of incorporating lost data eg. phase into the encoded data that is presented to the model.

1.4 Evaluation Methods

It shall be necessary for the project to set out a set of criteria for carrying out literary reviews on existing Learning Based Sound Synthesis Models. This must be done to decide if any method has potential to be built on further.

- Overly complex methods should be penalised. These can come in many forms eg. excessive training and computation time, extremely large datasets.

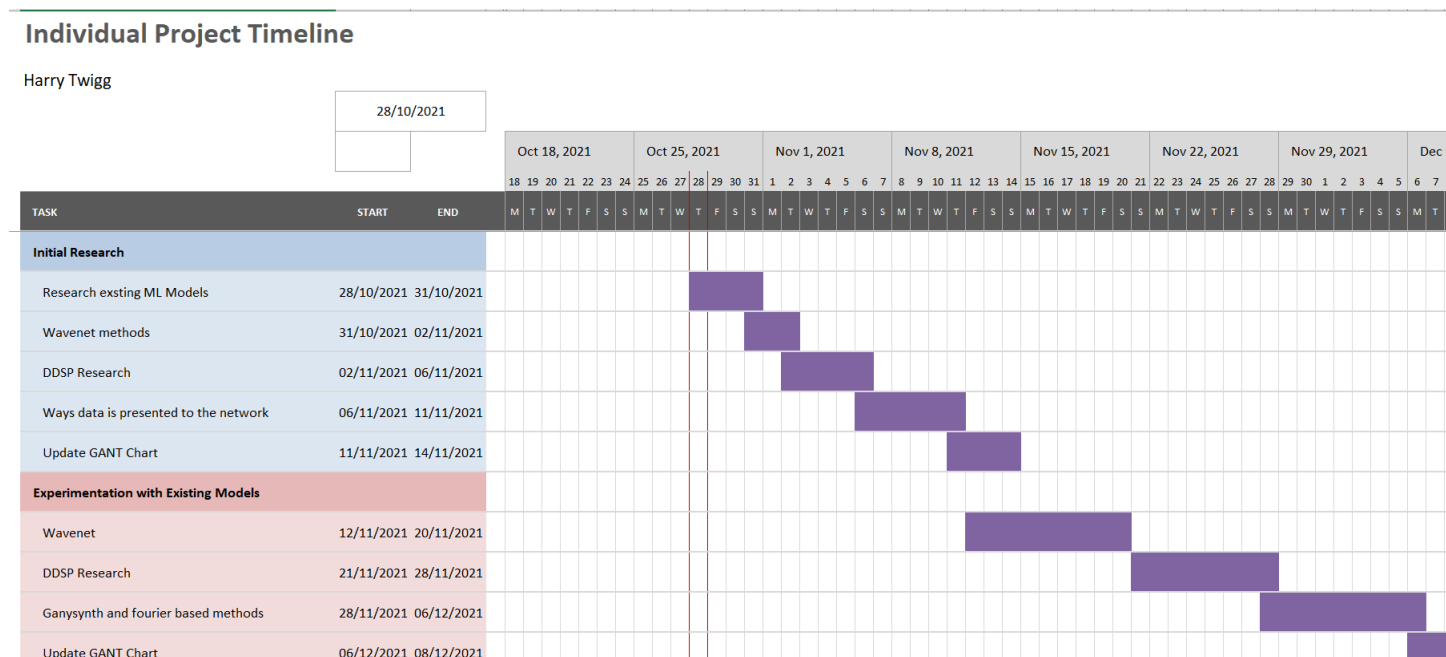
- Use of teacher forcing in any methods, as this will lead to biases in the model outputs.
- Poor tonal quality in the output, eg. it is noticeable that the model was generated digitally as opposed to recorded. This could be caused by
 - Spectral leakage due to inaccuracies in fourier representations
 - Poor oscillatory output representation that does not sound natural
- Modular systems shall be evaluated positively, due to the fact that their individual elements can be built on separately, and the whole system acts less like a 'Black Box'.
- Any discarded information eg phase that has been discarded during encoding (eg. phase) that could be presented to the network.

Chapter 2

Timeline

The project timeline is likely subject to change, so an adaptive plan will be used. Some parts will likely take longer than planned, and some parts will take less, this is incorporated into the plan. An initial estimate is given in the Gantt chart below.

For the full Gantt chart see the Associated Spreadsheet.



Chapter 3

Other Requirements

3.1 Training Needs

Extensive knowledge into how Machine Learning Models work, especially in relation to audio signal processing is essential for this projects progression. This shall be picked up in the initial research phase.

Being able to efficiently and accurately conduct a literary review is a skill that must be applied especially in the early phases of the project. Further knowledge of how to do this may be required.

Knowledge of how to access computing services shall be required potentially as the project progresses (See Resource Access).

3.2 Risk Assessment

No specific risk assessments shall be required for this project, no access to labs or other specialist equipment is required. The entirety of the project shall take place digitally.

3.3 Resource Access

No purchasing of physical resources for the project shall be required. However, it is likely that as the project progresses, access to Machine Learning optimised computer systems shall be required (eg graphics processing units or tensor processing units). This may take the form of access to the University's computing cluster.

Access to datasets eg. of preclassified information to feed into any machine learning models shall be required.