# Generating Experimental Audio With LSTM for Video Game Music

Ashwathi Pillai
20029015

21/22 MSc Data Science and AI for the Creative Industries

University of the Arts London

Unit Title: Artificial Intelligence for Media

Unit Code: IU000133

Assignment : Element 2 (Assignment 3 – Audio Generation Report)

Course Leader: Dr. Louis McCallum

Unit Tutors: Dr. Louis McCallum, Josh Murr, Vítek Růžička

16/3/2022 Word Count: 850

**Aim:** To create a model that generates new and interesting audio from a musical input for background scores in video games.

**Dataset For Training:** Three training inputs were utilised separately for this project:

1) Firstly, an audio sample from Björk's Tabula Rasa music video on YouTube was used to train the model.

2) To explore whether the generated output would be slightly different if an upbeat audio was utilised, a snippet from Björk's Earth Intruders music video on YouTube was used.

3) Finally, for the third training input, audio snippets from these three YouTube videos were used to create one single audio file for the model:

   a) Singing Bowl + Water

   b) Björk Talking About Her TV

   c) A Mind-Blowing Sitar Player

**Process:** A YouTube Downloader was used to extract sound from music videos. This project focuses on Björk's work for consistency. Tabula Rasa was picked as the first input to see how the model would generate audio using a soft input. It was trained on 500 epochs, without any customisation to the sequence array. The output generated can be found in the outputs folder in this repository, audio file named as 'output 1'

The generated audio sounds very different from the input file, although around 3-5 seconds, certain components of the song can be recognised if one is aware of what song was used for training the model. Then, a bit of tweaking was done to the sequence array to see if there were any interesting changes. After training on 500 epochs, the three arrays, just like the default code, had the following start points and segment length:

1) Starts generating 30% through the song and generates 400 frames

2) Moves to 60% through the song and generates 600 frames

3) Moves to 90% through the song and generates 150 frames

The output generated can be found in the outputs folder in this repository, audio file named as 'output 2'

The first output sounded a bit more louder than the second one, and despite the audio sounding quite muffled – it had some melody to it. The second output after experimentation sounded a bit bland.

For the second part, a more upbeat song by Björk titled Earth Intruders was picked. This time the model was trained on only 300 epochs (due to issues with Colab) with the following changes to the sequence array:

1) Starts generating 10% through the song and generates 2000 frames
2) Moves to 60% through the song and generates 500 frames
3) Moves to 90% through the song and generates 150 frames

The output generated can be found in the outputs folder in this repository, audio file named as ‘output 3’

A longer output was generated due to increased frames. Though it’s hard to tell if there are any similarities to the input audio, this output sounds a lot more suitable audio to work with for background scores for video games.

Finally, to verify if the reason why the output from the second audio input sounded more balanced, an audio file of soft sounds such as sitar, singing bowl, and some speech from a Björk clip was manually edited. The model was trained on 300 epochs with the following changes to the sequence array:

1) Starts generating 10% through the song and generates 200 frames
2) Moves to 30% through the song and generates 350 frames
3) Moves to 60% through the song and generates 400 frames
4) Moves to 90% through the song and generates 150 frames

The output generated can be found in the outputs folder in this repository, audio file named as ‘output 4’

The generated audio seems to have properly picked up on the speech snippet, as opposed to the sitar and singing bowl bits (despite all of them being equally present in the audio file). It is unclear as to why this has happened.

**Reflections:** Audio production for video games is a very exciting area with a lot of potential for creative exploration for various details i.e. a character sensing danger, browsing various weapons / avatar changes, new character introduction, etc. Therefore, it’s safe to assume that the person working with audio for games is contributing a lot to make the video game truly ‘alive’, as video game music is a very important element to the gaming experience. However, for indie game developers who do not have the budget to hire audio producers, or who do not have the

technical skillset to make their own music - generative techniques stand out as a great tool to encourage aspiring game developers to start working on their ideas without any limitations holding them back, thereby democratising the playing field. These techniques shouldn’t “replace” artists, but offer them a simple tool to give them more creative control while sampling. The model in this project generated some interesting outputs, but having more control over certain elements like pitch and pulse depending on the mood of the desired output would be ideal. For future iterations, perhaps using sequences created in Ableton as audio inputs could be done to see if there are any noticeable changes.