

---

# Project Proposal

---

## Group 11

Manas Angalakuduti, Matthew Arnold, Ryan Cooper, Pranav Kandarpa, Bradford Peterson

## Abstract

Generating music can seem uncomplicated when there are given parameters such as rhythm, dynamics, intensity, and pitch; however, generating music from an emoji or a simple one-word expression can be quite difficult. There have been many projects with similar goals, such as inputting a database of song lyrics and generating an accurate tag or genre for the song. We want to be able to generate music given an emoji, or mood, based on other songs with a respective 'emotional tag' that we would have the algorithm train. By training our algorithm using a database of songs' and their emotional tags, our goal is to generate a song given an emotional tag. This goal of producing music with machine learning methods will be achieved using a recurrent neural network.

## 1 Introduction/Background

Music generation has been attempted using neural networks since 1989, but it has gained traction with the discovery of deep neural networks' ability to learn from big data sets. As machine learning algorithms have become more widespread, and music streaming has become a larger part of people's lives, companies like Spotify and Apple have tried their hands at creating the most personalized listening experience. Spotify has recently taken a step forward in their efforts, and has patented software that could allow them to analyze user voices and recommend songs based on emotional state, accent, and other user data points. Similar music generation projects using neural networks have used long short-term memory recurrent neural networks to generate new music from previous training data. Likewise, we believe that using a Long Short-term Memory network will allow us to create pieces that resonate with people as appealing to a certain mood or emotional state.

## 2 Problem Definition

The inputs of this project will be a label and a corresponding MIDI dataset that will describe a mood. The label will describe each dataset as an emotion, described verbally or through an emoji. These inputs will be predefined, and each label will accompany a data set that contains a multitude of melodies with the given label. After inputting this into the neural network, the goal of this project is to output a melody that successfully replicates the mood described. We plan to use a long short-term recurrent neural network to generate a novel MIDI sequence given a data set of MIDI sequences depicting each mood. Because the perception of emotions in music may vary from person to person, certain assumptions must be made. For example, the mood for each emotion-based MIDI dataset will be predetermined and will be labeled based on the group's opinion of songs.

## 3 Methods

We will be using a deep recurrent neural network to generate a new MIDI sequence given training data, which will be predetermined MIDI sequences. We will be using MIDI versions of commonly used songs for each emotion (i.e. "Happy" by Pharrell for training a new happy MIDI sequence or

“Someone Like You” by Adele for training a new sad MIDI sequence). Researchers in the past have used a Long Short-Term Memory (LSTM) recurrent neural network to generate new music based off of training music data. The novelty of this project is classifying the emotional state of the music. Rather than creating new music based off of a composer or a genre, we hope to analyze data across various genres to create new music that invokes a certain emotion.

## 4 Metrics

Given a set of midi notes  $X_t, X_{t-1}, \dots, X_0$ , representing vectors at the time intervals  $t, t-1, \dots, 0$ , we would like to generate the most likely vector  $X_{t+1}$  representing the midi notes at the next time interval  $t+1$ . Because we are trying to predict the next  $t+1$  vector based from the previous data, this is a regression problem, and the loss function can be modeled as the L2 loss function:

$$l(\theta) = \frac{1}{T} \sum_t (X_t - \hat{X}_t)^2$$

## 5 Potential Results

As stated in the metrics section, our goal would be to minimize our L2 loss function. A similar project to ours, GRUV: Algorithmic Music Generation using Recurrent Neural Networks (Nayebi, Vitella) also used an L2 loss function to quantify the errors of their neural network. Likewise, in our project, the performance of the potential results will be displayed by a graph with the number of epochs on the x-axis, and LSTM loss on the y-axis.

In addition to a graph depicting the loss function over time, we obviously hope to generate music from our neural network. The length of the generated music has not been determined yet, however, we will be generating a new MIDI file that contains the newly generated notes. This MIDI file could be converted into audio using a MIDI editor or a digital audio workstation. Our group can qualitatively discern the quality of the music, and decide if it is pleasing to the ear. Overall, our potential results will include a graph indicating the performance metrics of our algorithm, and they will include MIDI files of newly generated music.

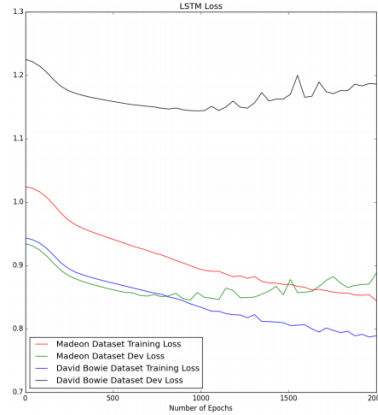


Figure 1: Example of loss function

## 6 Conclusion

This project will teach us about the patterns in music that lead humans to perceive specific emotions. By replicating and measuring these metrics of emotion in audio we can learn more about the principles of musical composition, and ideally create a model capable of generating music that can fit the emotion of a user's choice. In the evaluation of the model, the project also will most likely contribute more metrics or methods that can be used to measure the classification of emotion in music.

## References

- [1] Briot, Jean-Pierre, et al. Deep Learning Techniques for Music Generation. Springer, 2020.
- [2] L. Casini, G. Marfia and M. Rocchetti, "Some Reflections on the Potential and Limitations of Deep Learning for Automated Music Generation," 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna, Italy, 2018, pp. 27-31, doi: 10.1109/PIMRC.2018.8581038.
- [3] Conklin, Darrell, Music Generation from Statistical Models, 2003, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.3.2086>.
- [4] Cruz, Ricardo, et al. I-Sounds Emotion-Based Music Generation for Virtual Environments, 2007, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.101.2517>.
- [5] Nayebi, Aran, and Matt Vitelli. GRUV: Algorithmic Music Generation Using Recurrent Neural Networks, 2015, [cs224d.stanford.edu/reports/NayebiAran.pdf](https://cs224d.stanford.edu/reports/NayebiAran.pdf).
- [6] Yang, Li-Chia, et al. MidiNet: A Convolutional Generative Adversarial Network for Symbolic-Domain Music Generation, Cornell University, 2017, [arxiv.org/abs/1703.10847](https://arxiv.org/abs/1703.10847).