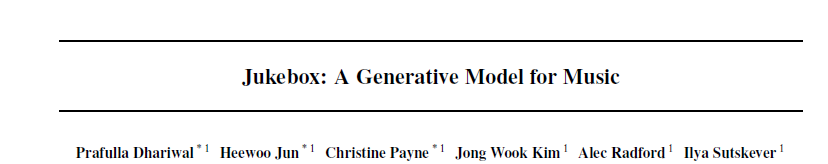
**Automatic Music Generation using AI**

1)Framing the Question

a)Goal-To generate a new piece of music using part of a song(By taking raw audio and encoding it and after implementing a model, do their generation process on this smaller variable set and decompress back to the original number of variables

b)Literature Review-

i)



This paper introduces a self made model Jutebox that generates music by using a multiscale

VQ-VAE to compress it to discrete codes, and modeling those using autoregressive Transformers.

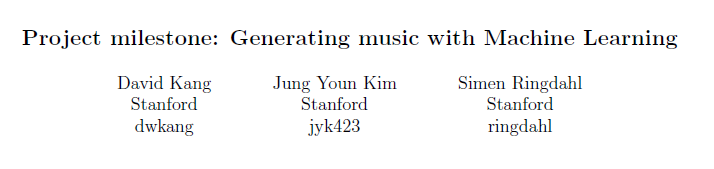
Training Dataset-44 kHz audio is compressed in dimensionality by 8x, 32x, and

128x respectively, with a codebook size of 2048 for each

level.

Model- The VQ-VAE has 2 million parameters and is trained on 9-second audio clips on 256 V100 for 3 days.

ii)



This report approaches music generation in four ways: one through a simple Naive Bayes algorithm and

the others through neural networks, specially a vanilla neural network, an LSTM RNN, and an encoder-decoder model RNN. For each algorithm, we will utilize different

approaches to data organization and music creation

the dataset would be represented by the vector [`C5', 1.0, `B-3.D4', 2.0], where the first two entries refer to the notes played by the right hand and their duration while the other two entries refer to that

of the left hand.

Code link- <https://www.dropbox.com/sh/ttzb502hheo9fst/AACx3uQSE_CSxnyp6bvaMQzZa?dl=0>

c)Ingredients-

i) Representations of songs are encoded in such a way each component of a vector represents it’s properties(instrument,frequency,chord,quality,shrillness,base,timeframe,etc).These will be the inputs contained in variables and the outputs will be the the predicted encoded representations of the next few seconds of the song which can be decoded based on the properties mentioned.

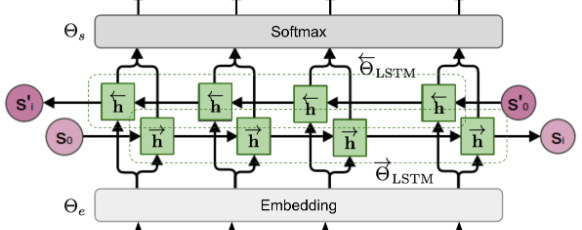
d)Quantifying hypothesis-Can we predict the output of a representation of a music which on decoding produces a note of music similar to that made by a musician.

2)Implementing the model

## a)Toolkit-Python(Modules to be imported-Numpy, Pandas, Pytorch, keras , Music21)

b) Model- Bidirectional LSTM layers that takes a matrix as an input and can returns a matrix, Dropout layers to prevent overfitting, Dense layers with Relu and softmax activation, where each input node is connected to each output node , Flattening and Embedding layers.The loss function would be categorical crossentropy ,with an optimizer(adam optimization),**The Activation layer** determines what activation function our neural network will use to calculate the output of a node.

Output which has to be decoded



Input- music in form of embedded vectors representing various

features

## 3)Model Testing-The model can be tested by running it on a previous song where the part of song to be predicted can be encoded as a testing set and every vector can be compared .Hence an accuracy for the model can be evaluated based on which further changes can be made