

# Generation of Music using LSTM Recurrent Neural Networks

[1]  
Sampoorna Agrawal  
School of Computer Science  
and Engineering  
Galgotias University  
Greater Noida, India  
sampoornaagrawal4@gmail.com

[2]  
Rajat Jain  
School of Computer Science  
and Engineering  
Galgotias University  
Greater Noida, India  
rajatjain.em@gmail.com

[3]  
Chirag Khurana  
School of Computer Science  
and Engineering  
Galgotias University  
Greater Noida, India  
chirag.rock1998@gmail.com

*Abstract - AI calculations have progressively been carried out in an assortment of ways inside the music area. In this paper, a Long Short Term Memory Network (LSTM) was utilized to tackle the music-related issue of arrangement. The organization was prepared on "Ryan's Mammoth Assortment of Fiddle Tunes" which was gotten from the music21 corpus. Past endeavors to produce tunes utilizing neural organizations have experienced an absence of "worldwide coherency". All organizations had the option to learn key. The 256-time step network delivered tunes that were the most worldwide intelligent, anyway the PC produced tunes were not decided to be as worldwide intelligent as the tunes in the preparation set.*

## 1. Introduction

Researchers have for some time been keen on utilizing AI strategies to create unique tunes. This assignment is deceptively troublesome. The decision of music portrayal plot, the quality what's more, amount of the information the organization is prepared on, and characterizing "great" tunes are largely non-paltry parts of the music-producing framework which should be considered during the plan cycle.

This paper will detail one such music-producing framework. Most examination on PC created tunes utilizes repetitive neural organizations (RNNs), since they are explicitly intended to "examine criticisms units and defer administrators, which permit the consolidation of nonlinearity and dynamical angles to the model" [1]. This implies that RNNs are particularly acceptable techniques for learning fleeting connections in successions. Mozer built up the Show engineering, an Elman-style intermittent neural organization to create unique tunes in view of the class of preparing information. His framework changed the entirety of its preparation information to

be in one key also, utilized an intricate portrayal conspire including family members pitches. Mozer had the option to produce songs that sounded charming to credulous audience members. Nonetheless, he didn't track down that the organization was ready to learn "more worldwide expression structure" [4]. Or maybe it was restricted to just initiating nearby connections among notes.

Later work kept on tending to this issue of "worldwide intelligibility": Eck and Schmid Bauer proposed the utilization of a particular kind of intermittent neural organization, a LSTM, since "Long Short Term Memory (LSTM) has prevailing in comparative areas where other RNNs have fizzled, like planning and then some, and learning of setting delicate dialects" [3]. This is on the grounds that LSTMs are explicitly intended to manage the issue of long haul conditions, which standard RNNs couldn't to address. Their music portrayal conspire was additionally far more straightforward than Moser's, basically allocating one info/target unit per note, with a 1 addressing on, and a 0 addressing off.

Not exclusively was this portrayal less complex, it likewise permitted there to be no fake differentiation among notes and harmonies. Eck and Schmid Bauer had the option to show that an "LSTM can play the blues with great planning and legitimate construction up to one will tune in" [3].

Later work in music piece has presented "turbulent motivation" to LSTM-based frameworks, which includes giving the organization a turbulent tune as "introductory motivation" [1]. Analysts have likewise endeavored to characterize "pleasantness" in some kind of quantifiable way, notwithstanding the abstract idea of melodic "quality". The framework proposed

in this paper will utilize a LSTM organization, a straightforward paired music portrayal plot, a specific measure of arbitrariness while producing songs, and a semi-abstract proportion of "worldwide coherency" which depends on human appointed authorities, to address a portion of the issues in the field.

## 2. Methods

Preparing information was acquired from the music21 corpus and preprocessed utilizing music21's object-oriented tool compartment. Three LSTM networks shifting in input-size were prepared on this information and afterward used to create five unique tunes each.

### 2.1 music21

music21 is an object-oriented python toolbox for looking, investigating, and controlling music scores. Scores are addressed as stream objects which can be settled and contain notes, rests, also, other representative information. All components are planned, in light of counterbalance from a beginning stage. music21 additionally accompanies a huge virtual corpus containing numerous works of old-style music, just as a huge assortment of society melodies [2]. For the motivations behind this undertaking, "Ryan's Mammoth Collection of Fiddle Tunes" - containing more than 1,000 songs - was picked since it was the biggest single-sort assortment in the corpus. Ryan's contained tunes in changed keys which were kept with no guarantees and not changed over to one single key.

### 2.2 Music Representation

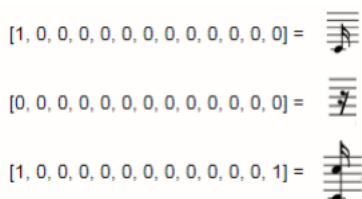


Figure 1 : Each array represents the melody at a certain offset.

Every tune in Ryan's assortment was changed into a grouping of exhibits. This was finished utilizing the music21 techniques Flatten and Notes and Rests which decreased each score to a rundown of note and rest components, each doled out a counterbalance esteem. The strategy GetElementsAtOffset was then used to check what components were playing at every sixteenth note. The notes and rests at each

counterbalance were then encoded into a cluster with 13 passages, every one of which addresses note from C4 to C5. Assuming a section is set to 1, that note is playing at that counterbalance. In the event that it is set to 0, isn't playing at that balance. This portrayal plot is very basic. Harmonies can without much of a stretch be addressed by essentially setting more than one section in a cluster to 1. Notwithstanding, the greatest downside of addressing tunes this way is that it's absolutely impossible to recognize rehashing sixteenth notes and notes of longer span. This downside was disregarded for time, despite the fact that enhancements to portrayal will be talked about further in resulting areas. Figure 1 shows instances of this portrayal. Every tune was cushioned with a specific measure of rests toward the start all together to offer the organization a chance to figure out how to start songs too.

### 2.3 LSTM

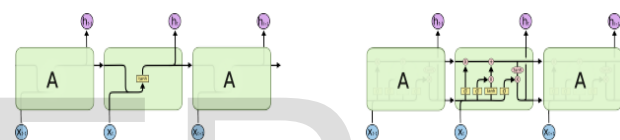


Figure 2: Left: The repeating module in the standard RNN contains a single layer. Right: Repeating module in a LSTM contains four interacting layers.

RNNs address the impediments of standard feed-forward models by including loops. This permits data to persevere. RNNs' chain-like construction makes it an especially appropriate model for managing with lists and sequences. Anyway, standard RNN models couldn't deal with long haul conditions, or when important data isn't later in the grouping. This is the place where LSTMs come in.

LSTMs were explicitly intended to address this weakness of RNNs. As opposed to the single layer module utilized in standard RNNs, LSTMs have four layers in every module. These layers permit each square of the LSTM to add, eliminate, or pass on data from its memory. LSTMs have demonstrated to be massively effective at an assortment of undertakings and have advocated RNNs as a dependable technique for learning sequences.

## 2.4 Training

The keras profound learning library was utilized to develop the organization structures, explicitly the LSTM layer and the consolidation layer. Three LSTM networks were prepared on a similar preparing information.

These organizations just varied in the quantity of time-ventures into the past the organization considered. 64-time steps relate to one 4/4 measure, so 64-, 128-, and 256-time step input sizes were chosen as boundaries to the organization (comparing to 1, 2, and 4 measures). Something else, all boundaries were set to the default boundaries as controlled by keras. Each organization was prepared for 10 ages with a clump size of 32, accomplishing a precision pace of 45%, 47.5%, and 43% separately. With regards to music piece, the low exactness rate ought not be cause for stress. Since we need the organization to produce unique tunes, we need the organization to learn general principles for music piece as opposed to figure out how to anticipate the songs inside the preparation information consummately.

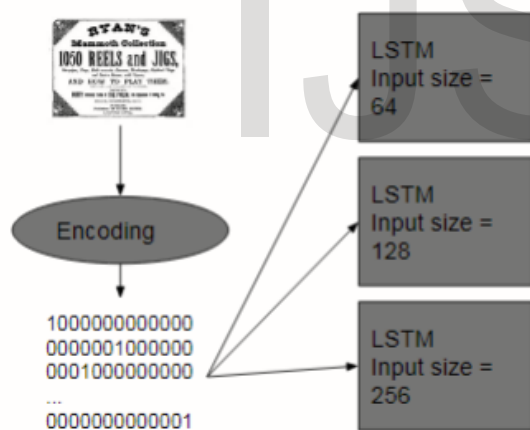


Figure 3: Three LSTMs were trained on encoded versions of melodies from Ryan’s Mammoth Collection of Fiddle Tunes.

## 3. Experiments

When each of the three organizations were prepared, the music-age measure started. Each organization was given five arrangements of a beginning information comprising of just rests. This was gone through the organizations which produced a likelihood appropriation for each conceivable next note. The following note was picked not by picking the note with the most elevated likelihood, but

instead from the likelihood dispersion. This permitted the created tunes to have a specific measure of controlled "turmoil". Songs regularly contain designs that are upset in astounding and unique manners, so by joining irregularity into the music-age measure some part of the imaginative interaction was additionally being demonstrated.

To address an issue referenced before in Section 2.2, whenever tunes were produced, rehashing notes were treated as the longest conceivable span except if there was an action break, in which case the notes were part separated at the action break. The topic of when to end the created songs was tended to just by subjectively picking a grouping length of 8 measures. This was educated by the way that a large portion of the fiddle tunes in the preparation information came in products of 4 measures, with 8 estimates being very normal.

Each produced tune was shipped off a group of two music scholars, Tai Warner and Asher Wolf, for investigation (alongside five songs from the preparation set). Examiners were uninformed of which tunes were produced by which organization. Every investigator delivered a worldwide coherency score from 1, not internationally intelligent by any means, to 10, consummately universally rational, in light of their examination of fiddle music hypothesis. Generally, worldwide soundness was decided to mean rehashing examples and melodic structures/topics, just as harmony movements that bode well given fiddle tune shows.

## 4. Results

Analysts were able to identify the training data melodies (model in Figure4), giving them a normal worldwide coherency score of 9.8. Despite the fact that there was no critical distinction between input-sizes of 64 and 128, 256-time steps prompted somewhat more around the world lucid tunes than the other two organizations. Generally, the songs created were middlingly worldwide rational. Be that as it may, some fascinating highlights of specific songs were noted by investigators.

Time Steps	Average Global Coherence
64	6.2
128	6.0
256	6.6
Training Melodies	9.8

Table 1: Average Global Coherence Scores for Each Network (based on 5 melodies each).

Figure 5 shows one tune which was marked the experts "top choice", notwithstanding being given a low worldwide coherency score of 3. In spite of the fact that there is by all accounts no reiteration of designs or topics inside the tune, experts noticed that the tune helped them to remember the style of Bela Bartok, a twentieth century writer who was impacted by Hungarian people music. Figure 6 shows another illustration of a song given a low worldwide coherency score with the examiners noticing its peculiarly balanced beats. Figure 7 shows an illustration of a song evaluated exceptionally in worldwide rationality. It displays a comparative cadenced design on the two lines, demonstrating reiteration of a melodic thought. Each produced song held a solitary key all through. This demonstrates that the organizations had the option to learn key: the most fundamental component of worldwide coherency.



Figure 4: A melody given a global coherency score of 10. Music theorists easily identified melodies from the training set.

### 5. Discussion

The organizations' capacity to learn key demonstrates that LSTMs show guarantee for learning highlights of all around the world lucid songs. Positively the accessible writing shows that this is valid. Albeit worldwide coherency is by all accounts inexactly corresponded with LSTM input-size, more analyses are expected to make any solid decisions about the information. The PC produced songs were effectively recognizable from tunes in the preparation information, demonstrating that there is still significantly more work to be done on the framework.

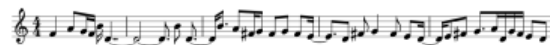


Figure 5: A melody given a global coherency score of 3, but labelled the music theorists' "favourite".



Figure 6: A melody given a global coherency score of 3. Music theorists' noted its weirdly offset beats.



Figure 7: A melody given a global coherency score of 8. Music theorists' only noted that the melody did not resolve at the end.

Be that as it may, in spite of some disillusioning outcomes, reactions from the experts towards the computer-generated tunes were very fascinating all alone, independent of a song's worldwide coherency. Truth be told, Warner and Wolf discovered the absolute least around the world lucid tunes of the bundle most fascinating. This recommends that there is some an incentive to a framework which doesn't accomplish "global coherency", yet in any case creates fascinating and unique tunes. Especially on account of "writers block", an organization that can produce sudden songs could start startling thoughts in authors.

Tracking down a quantitative method of assessing the nature of produced songs was troublesome. The privileging of Western music hypothesis and its accentuation on tune is one approach to doing as such, and truth be told is the means by which I and most work in the field has moved toward the assignment of PC music organization. Nonetheless, this is a restricted comprehension of music and what it is about music that energizes us to such an extent. I'm especially intrigued by test music from around the globe - music which expressly defies guidelines yet by the by can charm audience members. Could a particularly trial, rule-breaking, machine at any point exist?

## 6. References :

[1] Liang Zhao Andres E. Coca, Debora C. Corra.  
Computer-aided music composition with lstm neural  
network and chaotic inspiration.

*Neural Networks (IJCNN)* , 2013.

[2] Michael Scott Cuthbert nd Christopher Ariza.  
Music21: A toolkit for computer-aided musicology  
and symbolic music data. *ISMR*, 2010.

[3] D.Eck and J.Schmidhuber. Finding temporal  
structure in music: blues improvisation with lstm  
recurrent networks. *Neural Network for Signal  
Processing*, 2002.

[4] Michael C.Mozer. Connectionist music  
composition based on melodic, stylistic, and  
psychophysical constraints. *Music and  
Connectionism*, 1992.

IJSER