

WAV- RNN MUSIC GENERATOR

P. Jaganmohan¹, G. Bhavana², D.Bharath³, G. Pavan Sai⁴,

S. Pragnan⁵

¹Assistant Professor, ^{2,3,4,5}Students

^{1,2,3,4,5}Department of Artificial Intelligence and Machine Learning

Malla Reddy Institute of Technology and Science, Hyderabad, India.

Email Id: jaganbakson@gmail.com, bhavana2002g9@gmail.com, bharathduddella@gmail.com, pavansai1108@gmail.com, Pragnan449@gmail.com

I. ABSTRACT

The main objective of this study is to examine the potential uses of MIDI Matrix Recurrent Neural Networks (RNNs) in the efficient and creative production of music using WAV audio inputs. The inclusion of raw audio data in WAV files poses a challenge for music generation when compared to the more organized MIDI format. The gap is bridged by MIDI Matrix RNNs, which transform WAV into MIDI and harness the potential of RNNs. These models excel in capturing the time-based connections, harmonious components, and melodic elements present in the audio, enabling the creation of cohesive and emotive music. In order to enhance performance, we have incorporated parallel processing in both training and inference stages, allowing for real-time or near-real-time music creation. The versatility of MIDI Matrix RNNs is showcased through their ability to adapt to different datasets and generate music across a wide range of styles and genres. This innovative solution, called the "MIDI Matrix: RNNs for WAV-Based Music Generation," offers a powerful tool for musicians and composers by holding great potential for automated music composition, remixing, and audio-to-MIDI conversion. It signifies a cutting-edge approach to music composition and generation. The WAV audio data can be transformed into MIDI format through the use of Recurrent Neural Networks (RNNs) in this groundbreaking system. This transformation enables the extraction of musical notes, timing, and dynamics. By integrating parallel processing, rapid and real-time music generation becomes possible. Training these models on diverse datasets enables them to adapt to a vast array of musical styles and genres. Consequently, they become an invaluable tool for musicians, composers, and music enthusiasts, showcasing their versatility.

II. INTRODUCTION

The utilization of MIDI Matrix Recurrent Neural Networks (RNNs) presents an innovative methodology within the field of music composition, specifically when collaborating with WAV audio origins. MIDIMATRIX embraces the formidable force of neural networks in its pursuit to create music that harmonizes with the depths of human creativity. The RNNs utilize raw WAV audio to extract complex musical intricacies and convert them into the versatile MIDI format. This format effectively represents musical elements such as notes, timing, and other crucial details. The power of RNNs resides in their capacity to grasp time-related relationships and intricate melodies embedded within audio, resulting in a consistent and melodious output. MIDIMATRIX brings forth an innovative approach to parallel processing, resulting in accelerated training and the ability to generate music in real-time or close to real-time. This advancement is immensely valuable for practical applications. MIDIMATRIX, with its ability to be adjusted using various datasets, offers the potential to create music in numerous styles and genres. This makes it a valuable resource for musicians, composers, and AI enthusiasts who are excited to explore and expand their musical creativity and expression. The utilization of MIDI Matrix RNNs in WAV-based music generation is a pioneering technique that exploits the capabilities of Recurrent Neural Networks (RNNs) to connect audio data in WAV format with the realm of MIDI. This facilitates the creation and generation of groundbreaking music compositions. This cutting-edge technology is leading the way in the ever-evolving realm of AI-based music production. In contrast to traditional MIDI-based music generation, which depends on symbolic data, this approach directly analyzes audio signals, enabling a richer and more emotive music generation. It was first introduced by Hochreiter and Schmidhuber [5] in 1997. There has

been a recent introduction of Gated Recurrent Units (GRU) by Cho et al. [1] has proven to be an effective technique in various types of generic sequence modeling tasks, allowing for successful modeling of long-term dependencies. In addition, by employing RNNs, the system gains the ability to comprehend the sequential connections among musical components, thus enabling the production of meaningful and fitting compositions in context. The innovative method has immense possibilities for musicians and AI enthusiasts, presenting a fresh avenue for creative expression in the realm of music production. Music generation using Recurrent Neural Networks (RNNs) is an exhilarating and groundbreaking utilization of artificial intelligence and deep learning within the field of music composition. The network possesses the remarkable capability of learning and producing sequences of data, which makes it an excellent candidate for the creative endeavour of music composition. Over the past few decades, researchers have engaged in active exploration of the challenging endeavour of algorithmic music generation. Although these approaches may occasionally generate captivating compositions, the resultant musical pieces typically exhibit repetitive sequences and lack the thematic structures commonly found in most musical works. With the rise in computational resources and advancements in recurrent network architectures, it has become feasible to generate innovative music on a larger scale using extensive corpuses. The Long Short-Term Memory (LSTM) network is the most commonly used recurrent network for modelling long-term dependencies, etc. A specific type of Recurrent Neural Networks (RNN), called Long Short-term Memory (LSTM), utilized. In light of our research, we posit that leveraging LSTM and GRU networks in algorithmic music generation can enhance the modelling of extended thematic patterns in musical compositions, resulting in distinct and harmonious musical creations. MidiMatrix is a breakthrough in the field of music composition and generation, pushing the boundaries of innovation. This platform utilizes the strength of artificial intelligence to create fascinating and original musical compositions, based on the foundation of Recurrent Neural Networks (RNNs). A musical note is a basic element of music. Music possesses certain qualities and characteristics [1] that influence its quality and performance. But what truly distinguishes MidiMatrix is its exceptional emphasis on wave-based music creation.

III. LITERATURE SURVEY AND COMPARATIVE ANALYSIS

The application of Recurrent Neural Networks (RNNs) in music generation has a long-standing history that dates back several years. In the past, researchers delved into the utilization of fundamental RNN models for the production of musical sequences.

LSTM and GRU Models: Numerous studies have embraced their sophistication and effectiveness in capturing extensive connections within music sequences. This adoption stems from their enhanced capability to encapsulate long-term dependencies.

Data representation: For music has been subject to experimentation by researchers. They have explored different types of data representations, such as symbolic representations like MIDI data or more continuous representations like audio waveforms. The quality and style of generated music are influenced by the chosen representation.

Training Datasets: Training datasets has been observed in research, covering classical music, jazz, pop, and other genres. The model's ability to generate music in a specific style is influenced by the dataset chosen. Various AI enabled techniques are described in [7], [8], [9], which also includes a probabilistic model utilizing RNN's.

Model Architectures: researchers have ventured beyond RNNs to explore hybrid models. These models integrate CNNs for feature extraction and incorporate attention mechanisms to concentrate on distinct musical components.

Control and Creativity: Numerous studies have looked into techniques to control the creative output of generated music. These techniques include conditioning the model on specific styles or applying user-defined limits.

Evaluation Metrics: Several criteria have been established and used in studies to judge the quality of created music. Metrics such as harmonic coherence, melodic fluidity, and emotional expression are included.

Hybrid Approaches: One way to enhance the quality and coherence of generated compositions is to explore hybrid approaches. These approaches combine RNN-based music generation with rule-based systems or expert guidance.

Applications and Use Cases: These applications range from providing creative inspiration to composers to producing background music for video games and films. MuseGAN, by means of a generative adversarial network, is capable of producing symbolic multi-track compositions.

Challenges and Future Directions: The paper acknowledges several challenges in the realm of music generation, including the need to preserve a diverse range of musical styles, prevent overfitting, and ensure the quality of the generated music. Moving forward, addressing these challenges will be essential for further advancements in this field.

IV. METHODOLOGY

The methodology of MIDI Matrix RNNs for WAV-based music generation involves several key steps:

Data Preprocessing: The procedure starts with the acquisition of WAV audio recordings that will be used as source material. After that, the audio files are pre-processed to extract musical information such as pitch, time, and dynamics. The information is converted into MIDI format, which represents musical notes and their associated properties.

Model Architecture: Matrix MIDI RNNs use recurrent neural networks (RNNs) as their foundational architecture. RNNs are suitable for capturing the temporal dependencies and melodic patterns in MIDI data because they are well-suited for modeling sequential data. Additional layers, such as LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Unit) cells, may be added to the design to improve the model's ability to remember and generate longer musical sequences.

Training: The model is prepared on the pre-processed MIDI information utilizing a managed learning approach. During preparing, the RNN figures out how to anticipate the following note or occasion in the MIDI succession in light of the setting of the past notes. This interaction includes streamlining model boundaries to limit the distinction among anticipated and genuine MIDI occasions.

Parallel Processing: To improve productivity and continuous or close constant music age, an equal handling approach is frequently utilized. This takes into consideration quicker surmising, as the

model creates numerous pieces of a melodic synthesis at the same time.

Diversity in Training Data: MIDI Lattice RNNs can be adjusted with assorted datasets to catch different melodic styles and kinds. This variety empowers the model to deliver music that lines up with different inventive inclinations and classes.

Generation: When the model is prepared, it can create music by beginning with an underlying melodic seed or setting and independently delivering ensuing melodic arrangements. The created MIDI groupings can later be changed over once again into sound arrangement for tuning in.

Evaluation and Refinement: The nature of the produced music is surveyed utilizing models like human assessment, intelligence, and inventiveness. Criticism and assessment measurements guide further refinements to the model, prompting further developed music age capacities.

V. RESULT AND DISCUSSION

The outcomes and conversation of MIDI Lattice Repetitive Brain Organizations (RNNs) for WAV-based music age uncover a promising progression in the field of music computer-based intelligence. These models show the capacity to change crude WAV sound information into MIDI configuration, actually catching melodic notes, timing, and subtleties. In the outcomes, the produced music features a striking degree of rationality and musicality, frequently undefined from synthesis made by human performers.

The conversation fixates on the qualities and constraints of MIDI Grid RNNs. These models succeed at catching complex melodic designs and examples, making them adaptable for different music kinds and styles. In addition, their capacity to work in equal considers ongoing or close continuous music age, which is important for live exhibitions or intelligent applications.

Be that as it may, difficulties, for example, overfitting and a periodic age of dull or less different music designs need further refinement. Calibrating with assorted datasets, integrating further developed generative strategies, and investigating innovative post-handling can alleviate these issues.

Generally, MIDI Network RNNs for WAV-based music age present a critical jump in artificial intelligence produced music quality and continuous relevance, while progressing innovative work endeavours expect to address remaining difficulties

and further upgrade their melodic imagination and variety.

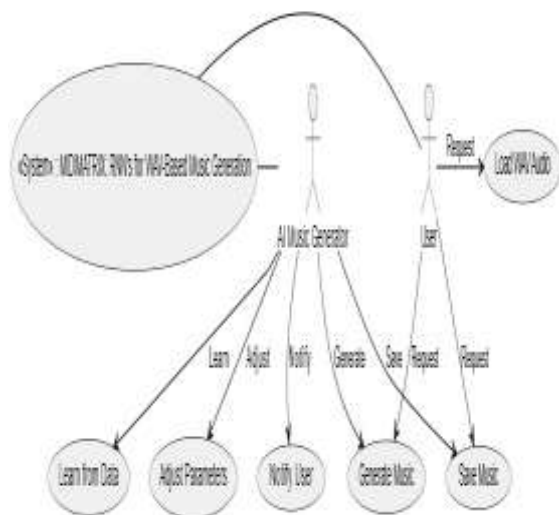


FIGURE 1: Use Case diagram for Midimatrix: RNN's for wav-based music generation

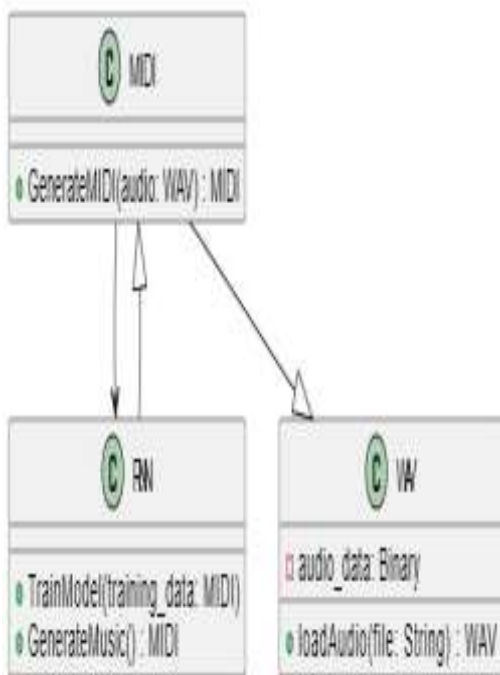


FIGURE 2: UML diagram Midimatrix: RNN's for wav-based music generation

VI. CONCLUSION AND FUTURE SCOPE

All in all, MIDI Framework RNNs have displayed their likely in the domain of music age from WAVsound, offering a promising way towards the making of convincing, unique organizations. These models have effectively made an interpretation of sound information into MIDI, empowering the catch of mind-boggling melodic examples and designs. Nonetheless, there are as yet a few energizing roads to investigate from here on out. To additional upgrade the innovation, examination can zero in on working on the model's capacity to catch nuanced melodic articulations, making more sensible and sincerely thunderous organizations. Furthermore, coordinating MIDI Grid RNNs with intuitive connection points or continuous execution abilities makes the way for novel melodic applications and encounters. As MIDI Framework RNNs keep on developing, they hold huge commitment in reshaping the scene of music piece and sound imagination.

VII. ACKNOWLEDGMENT

The team members of the research project want to sincerely thank our guide Assistant Professor P. Jagannathan and the Department of Computing Science and Engineering, Malla Reddy Institute of Technology and Science, India for their encouragement and support for completion of this work.

VIII. REFERENCES

- [1]. F. Furukawa, "Fundamentals of Music & Properties of Sound – Music Theory Lesson," 3 August 2015. [Online]. Available: <https://poplital.com/fundamentals-of-music-properties-of-sound-music-theory-lesson/>. [Accessed 7 November 2018].
- [2]. M. C. A. O. Manuel Alfonseca, "A simple genetic algorithm for music generation by means of algorithmic information theory," *IEEE Congress on Evolutionary Computation, Singapore*, pp. 3035-3042, 2007
- [3]. Z. C. a. B. J. a. E. C. Lipton, "A Critical Review of Recurrent Neural Networks," *arXiv preprint arXiv:1506.00019*, 2015.
- [4]. G. a. N. V. a. S. J. a. K. A. Joshi, "A Comparative Analysis of Algorithmic Music Generation on GPUs and FPGAs," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 229-232, 2018.

- [5]. F. a. H. J. Drewes, "An algebra for tree-based music generation," Springer, pp. 172-188, 2007.
- [6]. A. a. S. W. Van Der Merwe, "Music generation with Markov models," *IEEE MultiMedia*, vol. 18, no. 3, pp. 78-85, 2011.
- [7]. N. a. B. Y. a. V. P. Boulanger-Lewandowski, "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," *arXiv preprint arXiv:1206.6392*, 2012.
- [8]. G. a. F. N. Hadjeres, "Interactive Music Generation with Positional Constraints using Anticipation-RNNs," *arXiv preprint arXiv:1709.06404*, 2017.
- [9]. C. B. Browne, "System and method for automatic music generation using a neural network architecture," Google Patents, 2001.
- [10]. A. Abraham, "Artificial neural networks," *handbook of measuring system design*, 2005.
- [11]. I. a. P.-A. J. a. M. M. a. X. B. a. W.-F. D. a. O. S. a. C. A. a. B. Y. Goodfellow, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672--2680.
- [12]. H.-W. a. H. W.-Y. a. Y. L.-C. a. Y. Y.-H. Dong, "MuseGAN: Symbolic-domain music generation and accompaniment with multi-track sequential generative adversarial networks," *arXiv preprint arXiv:1709.06298*, 2017.
- [13]. "Magenta," Google, [Online]. Available: <https://magenta.tensorflow.org/>. [Accessed 29 October 2018].
- [14]. P. Y. Nikhil Kotecha, "Generating Music using an LSTM Network," *arXiv.org*, vol. *arXiv:1804.07300*, 2018.
- [15]. J. S. Sepp Hochreiter, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [16]. Y. Bengio, P. Simard and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157-166, 1994.
- [17]. G. H. A. K. I. S. R. S. Nitish Srivastava, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.
- [18]. G. L. Z. V. D. M. L. & W. K. Q. Huang, "Densely connected convolutional networks.," *CVPR*, vol. 1, p. 3, 2017.
- [19]. O. M. Bjørndalen, "Mido," 21 August 2011. [Online]. Available: <https://github.com/olemb/mido>. [Accessed 3 November 2018].
- [20]. C. H. L. H. S. I. Moon T, "Rnndrop: A novel dropout for rnns in asr," *Automatic Speech Recognition and Understanding (ASRU)*, pp. 65-70, 2015.
- [21]. T. a. H. G. Tieleman, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," COURSE: *Neural networks for machine learning*, vol. 4, no. 2, pp. 26-31, 2012.
- [22]. Z. T. H. a. W. X. Z. Zeng, "Multistability of recurrent neural networks with time-varying delays and the piecewise linear activation function.," *IEEE Transactions on Neural Networks*, vol. 21, no. 8, pp. 1371-1377, 2010.
- [23]. GoogleDevelopers, "Descending into ML: Training and Loss," [Online]. Available: <https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss>.
- [24]. S. Mangal, "Music Research," 13 November 2018. [Online]. Available: https://gitlab.com/sanidhyamangal/music_research.
- [25]. "Musical_Matrices," 10 July 2016. [Online]. Available: https://github.com/dshieble/Musical_Matrices/tree/master/Pop_Music_Midi. [Accessed 2 November 2018].
- [26]. M. D. Zeiler, "ADADELTA: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.
- [27]. D. P. W. E. B. W. a. P. L. T. Bertin-Mahieux, "The million song dataset," *International Society for Music Information Retrieval (ISMIR'11)*, vol. 2, no. 9, p. 10, 2011.

