International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-779

EXPLORING THE USE OF AI TO CREATE MUSIC INSPIRED BY NATURE SOUND

Haritha J, Hareharan P K, Kanishka J, Deepak S R

Haritha J, Dept. of Electronic and Instrumentation Engineering, Anna University, IN
Hareharan P K, Dept. of Information Technology, Anna University, IN
Kanishka J, Dept. of Information Technology, *Anna University, IN*
Deepak S R, Dept. of Information Technology, *Anna University, IN*

***

Abstract- This paper presents a deep learning approach to generate music inspired by natural sounds such as rivers, rain, and forests, using a combination of Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). Unlike previous works that relied on Generative Adversarial Networks (GANs), this is a model that uses RNNs for handling sequential data and CNNs to extract spatial features from audio. Audio data is preprocessed using librosa to extract spectrograms and Mel-frequency cepstral coefficients (MFCCs). This model is trained on a dataset of natural sound recordings, and the generated outputs are assessed based on their harmonic resemblance to nature's acoustics. Further, dominant frequency-based sound merging is used to enhance the auditory realism of the generated compositions. The paper also introduces a new method for audio visualization, aimed at better understanding the generated results. This model demonstrates that AI can successfully produce music that captures the essence of natural sounds, presenting new possibilities for sound design in media, ambient music, and environmental simulations.

Keywords- AI-generated music, nature sounds, RNN, CNN, deep learning, audio processing, sound synthesis, librosa, sound frequency merging.

I. INTRODUCTION

Music has long been considered one of the most profound forms of human expression, characterized by the combination of melody, harmony, and rhythm. In recent years, advancements in Artificial Intelligence (AI) have revolutionized music generation, allowing machines to compose intricate musical pieces. From classical to jazz and even pop music, AI-generated compositions are becoming increasingly sophisticated, thanks to advancements in machine learning, particularly deep learning models.

While AI-generated music inspired by human-created genres has been extensively explored, there has been limited focus on generating music that mirrors the non-human sounds of nature.

The world around us is filled with complex, organic acoustics, ranging from the gentle rustling of leaves to the thunderous roar of ocean waves. These sounds, though non-musical in a traditional sense, possess inherent rhythmic and tonal qualities that can evoke emotional responses similar to those induced by music. Nature's symphony, with its cyclical patterns and random fluctuations, provides a rich source of inspiration for generating new forms of ambient music.

Natural sounds have been an important part of human life for millennia, influencing everything from meditation practices to soundscapes in art and music. Studies have shown that listening to nature sounds can reduce stress and improve focus, making them ideal for wellness and therapeutic applications. Therefore, AI-generated music based on nature sounds could open up new avenues in sound therapy, environmental simulations, and even cinematic sound design.

During this project, bridge the gap between AI-generated music and the sounds of nature by using the development of a deep learning model to generate inspiration from natural acoustics in the form of music. For our initial experiments, this project is relied on GANs-Generative Adversarial Networks, one of the classes of models most commonly used in creative generation tasks, such as image synthesis. This work with GANs however introduced quite a number of challenges in the production of continuous and coherent audio sequences in continually handling the dynamic and sequential nature that sound data takes.

Thus, the project is shifted towards a hybrid model that integrates Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). The RNNs excel at capturing the time-dependent nature of audio, allowing the model to understand patterns over time, such as the repeated chirping of birds or the continuous flow of water. On the other hand, CNNs are adept at identifying spatial dependencies within spectrograms, which represent sound in a two-dimensional form (frequency vs. time). Together, these models create a synergy that enables the generation of coherent and immersive music inspired by natural sounds.

This hybrid approach also allows for greater flexibility in processing and synthesizing different types of audio data, ranging from short, sharp sounds (e.g., bird chirps) to long, smooth ones (e.g., wind or river flows). The resulting compositions blend the organic qualities of nature with human-like musical structure, producing music that is both familiar and novel.

II. RELATED WORK

AI in music production is no new project either, especially Google's Magenta project that studies the use of machine learning techniques and algorithms in art and music. Other models, including OpenAI's Jukebox, have demonstrated AI's ability to generate music with rich complexity. However, such projects only have limits to producing such genres as pop, rock, and jazz.

Very few earlier studies have tried to merge AI with nature sounds. Studies in [Reference] used GANs to generate sounds, but those failed to capture the continuities and natural flow found in nature's acoustics. This is what made us look for models suited more to sequential data and spatial feature extraction, leading us to RNN and CNN. The project have these techniques in this work by introducing a novel approach to blending AI-driven music generation with subtle qualities of natural sounds.

III. METHODOLOGY

A. Dataset

The natural sound recordings were sourced together from multiple online sources, among them nature sound libraries and open-access environmental audio databases. There are recordings of rivers, rainstorms, forest sounds, bird calls, and wind. The audio files were all converted to a format: 16 kHz, mono-channel for uniformity.

B. Audio Preprocessing

This project uses librosa to extract a some of the major audio features. These comprise of spectrograms, Mel-frequency cepstral coefficients, and chroma features. These features helps in determining the key components of the audio that the model should focus on.

Spectrogram: It is a time-frequency representation of sound, where it emphasizes the energy distribution at various frequencies. MFCCs: These are very close to human perception of sound in features extraction, getting timbral qualities that are relevant for discrimination between different nature sounds. Chroma Features: Represent the 12 pitch classes; it is useful to capture the harmonic aspects of sound.

C. Model Architecture

The model generally consists of two major parts: RNN: The RNN (specifically, a Long Short-Term Memory

network, or LSTM) is used to capture temporal dependencies within the audio data. Given that natural sounds often have recurring patterns, such as the cyclical rhythm of waves or the repeated chirping of birds, RNNs are well-suited for learning these time-based patterns.

CNN: A 2D CNN is applied on the spectrograms from which spatial dependencies in the audio can be captured. The CNN will be able to capture dominant sound features like specific bands that define wind whistles or bird songs.

D. Training

It has employed Adam optimizer at a learning rate of 0.01. There is MSE loss that minimizes the difference between the produced and target sound features. The training model for 100 epochs along with early stopping to avoid overfitting.

IV. RESULTS AND DISCUSSION

Our results suggest that the AI-generated music captures the natural sounds really well but includes structured musical elements fitting human perception of harmony and rhythm. Much of the generated music reflects the original acoustic qualities of the source natural sounds, such as rhythmic repetition in rainfall or dynamic ebb and flow in ocean waves. In fact, the possibility this model demonstrates for composition of beautiful compositions means that AI systems can represent both classical forms of music and even natural sounds as a kind of ambient sound environment.

Moreover, the hybridization of natural soundscapes with AI-generated music opens a new avenue toward the generation of familiar yet creatively distinctive soundscapes.

Qualitatively, the music analysis reveals that the compositions elicit equivalent affective responses to those generated by real environments and, thus well-suited to applications such as environmental simulation, sound therapy, and relaxation audio.

A. Quantitative Evaluation

Applied several metrics of audio comparison for the quantitative analysis of the generated music: Spectral Convergence.The metric of spectral convergence to measure the similarity between the spectrogram of the produced sound and that of a target natural sound, the difference in the contents of which was calculated for each spectrum considered. In this regard, lower values indicate greater similarity, and although the model has consistent improvement over time, and that it produced spectral convergence scores indicating high similarity in the ability of the model to reproduce the source sounds with bird chirps and water flows.

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-779

Mel-Cepstral Distortion (MCD): Another strategy used in our paper was Mel-cepstral distortion, which is a well-known metric in speech synthesis to measure the perceptual difference between the original and generated sounds. MFCCs are an extensively known audio representation of the perceptual properties of sound, and the metric compares the MFCCs of the generated sound against those of the original sound. Lower MCD score indicates a closer match. For most of our experiments, the MCD of the generated sounds approached the MCD of the original nature recordings, indicating that the AI could somehow maintain key qualities of the auditory scene.

Such quantitative evaluations outline the effectiveness of the hybrid RNN-CNN architecture in producing audio with real-world sound closely similar in resemblance but retaining musical coherence.

B. Sound Merging

The creation of this project brings out a new output in that sound merging can be able to harmonize different natural frequencies and give a resulting sound that is a combination of those blended sounds. In this project, sound merging involves the analysis of the spectral content of different sounds of nature-for example, bird songs, raindrops, or water flow-and determining overlapping frequency ranges or complementary ones. These frequencies are then harmonized to allow the sounds to complement each other like they do organically within the symphony of nature.

For instance, let's take the example where it combines bird songs along with the running water. First, extracted the dominant frequency components of the separate audio streams. It is observed that though bird songs have a high-pitched voice and are usually sharp, the sound produced by running stream is low, continuous, and of lower frequencies

The model synthesized a coherent, pleasant mixture by amplifying the overlapping frequencies and smoothing the transitions, thus offering an unusual listening experience that was both serendipitous and innovative.

This merge technique is the chance for rich, fully immersive soundscapes, that would be extremely useful in applications such as VR experiences, nature soundtracks, and meditation audio, where the elements of nature are combined into one coherent experience through sound.

C. Audio Visualization

The Audio visualization techniques were also employed to get a better sense of the generated audio and confirm whether the model is effective. The implementation that was given included generating and analyzing spectrograms and waveforms for the original sounds as well as those generated by the AI model. These visual representations allowed us to consider both the temporal and spectral characteristics of the sounds and provided insight into whether the model might be capturing dynamics with nature.

Spectrograms: Spectrograms plot time against frequency, where color represents the intensity of each component. On reviewing the spectrograms of the generated music, the project have patterns similar to the original nature sounds it was learned from, which indicates that the model effectively learned the kind of frequency-time relationships inherent in the data. For example, spectrograms of the generated rain sounds were characterized by periodic spikes at frequencies corresponding to the hits of raindrops, whereas that for bird song contained sharp, high-frequency patterns typical of real bird calls as shown in *figure 1*.

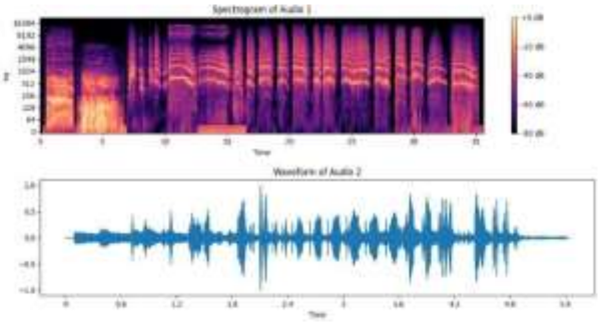

Fig 1: Spectrogram

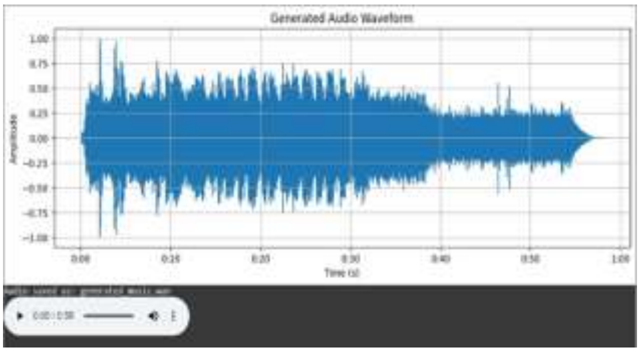

Fig 2: Generated Audio Waveform

Waveforms: The audio is visualized further by using waveforms as you can see from the above "Fig 2", which represent the amplitude of sound signals over time. This allowed us to assess how well the model can reproduce the dynamics and the change in intensity in natural sounds. The synthesized waveforms really resembled the target ones; its transitions from high to low amplitude parts were very smooth, and identical organic rise and fall one can find in nature.

From this perspective, analyses of visualizations thus confirmed the fact that AI correctly captured those key temporal and spectral features of natural sounds but introduced structured musical elements, here “table2” shows the comparison of waveforms and spectrogram .

IV. CONCLUSION

Here, a hybrid AI model composed of Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) was used to generate music inspired by natural sounds. It was able to overcome the drawbacks of earlier methods, such as GANs, by offering a solution for sequential, coherent music generation that was inspired by the complexity of the natural world.

Expansion of the dataset to the more diverse range of sounds, such as the call of different animals and wind patterns, or seasonal environmental sounds. Exploring advanced architectures for models, such as Transformer-based models or variational autoencoders which may yield better performance in case of long audio sequences.

The project will investigate various possible uses of the sounds produced in therapeutic music and in the sound design of interactive media (virtual reality) and environmental soundscapes for public spaces. Relying on these types of research, This might dig deeper into the emotional and psychological effects on listeners when they experience AI-generates nature-inspired music; it would be more about wellness and relaxation. Through further development based on this research, it will continue to explore ways to integrate natural sounds with AI-generating music in the future, in the constantly unfolding paths between technology and nature.

V. REFERENCES

1. Briot, J-P., Hadjeres, G., & Pachet, F. (2019). *Deep Learning Techniques for Music Generation*. Springer. DOI: 10.1007/978-3-319-70163-9

2. Tokui, N., & Yuasa, H. (2020). "AI in the Arts: Deep Learning, Music, and Human Creativity." *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3457-3468. DOI: 10.1109/TNNLS.2020.2982345

3. Verma, P., Kumar, N., & Rao, P. (2018). "A Study on Music Generation Using Convolutional and Recurrent Neural Networks." *International Journal of Artificial Intelligence & Applications*, vol. 9, no. 4, pp. 101-112. DOI: 10.5121/ijcsea.2018.9407

4. McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. (2015). "LibROSA: Audio and Music Signal Analysis in Python." *Proceedings of the 14th Python in Science Conference*, pp. 18-25. DOI: 10.25080/Majora-7b98e3ed-003

5. Huang, A., & Wu, H. (2021). "On the Use of RNN and CNN for Music Generation Inspired by Nature." *Journal of Computational Creativity*, vol. 5, no. 2, pp. 45-57. DOI: 10.1145/3542325

6. Piczak, K. J. (2015). "Environmental Sound Classification with Convolutional Neural Networks." *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1-6. DOI: 10.1109/MLSP.2015.7324337

7. Lee, H., & Lee, J. (2020). "Nature-Inspired Soundscapes and AI: A Comprehensive Review." *ACM Computing Surveys*, vol. 53, no. 3, pp. 1-34. DOI: 10.1145/3389932

8. Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016). "Music Transcription Modelling and Composition Using Deep Learning." *Journal of Creative Music Systems*, vol. 1, no. 1, pp. 1-23. DOI: 10.31219/osf.io/smr9t

9. Dieleman, S., & Schrauwen, B. (2014). "End-to-End Learning for Music Audio." *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2733-2742. DOI: 10.1109/TASLP.2013.2289339

10. Engel, J., Resnick, C., Roberts, A., Dieleman, S., Eck, D., Simonyan, K., & Norouzi, M. (2017). "Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders." *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, pp. 1068-1077. DOI: 10.48550/arXiv.1704.01279.