The Generation of Piano Music in the Style of Johannes Brahms Using Neural Network Architectures

James Doherty
Technological University Dublin
Central Quad, Grangegorman Lower, Dublin 7, D07ADY7
Dublin, Ireland
e-mail: d14126143@mytudublin.ie

Brendan Tierney
Technological University Dublin
Central Quad, Grangegorman Lower, Dublin 7, D07ADY7
Dublin, Ireland
e-mail: brendan.tierney@tudublin.ie

Abstract - **Neural network architectures currently are only able to employ music generation tasks to similar levels of human composers when the music is at a basic compositional standard as they struggle with the complex motifs and harmonic structures of Western Classical Music. This study aims to determine if various data preprocessing and augmentation techniques can train a neural network model to generate pieces of piano music to a similar level of musicality and emotion as Romantic Period composer Johannes Brahms. Quantitative experimentation involving Music Information Retrieval was conducted, as well as a quantitative survey with respondents consisting of only professional musicians, composers, and conductors. Analysis of the results demonstrated that Transformer models using various attention mechanisms generated statically similar results to the original piano works of Brahms and that survey participants struggled to distinguish between the pieces generated by Brahms and the models. The results indicate that various data preprocessing and augmentation methods do have an impact on model accuracy resulting in the ability to generate longer sequences of music containing the composite motifs and harmonic structures of romantic period piano music.**

Keywords-artificial Intelligence; music generation; neural network architecture; Brahms.

I. INTRODUCTION

The intention of this project is to generate piano music in the style of classical music composer Johannes Brahms by training a Recurrent Neural Network (RNN), a Long Short-Term Memory (LSTM) based RNN, Transformer models with different attention mechanisms, and a Perceiver AR model. These models will be trained with a pre-processed and augmented dataset containing MIDI files of Brahms' piano works. To deem the success of the project, the best musical pieces generated from the neural network models must show statistical similarities in various musical variables using Music Information Retrieval (MIR). Along with this, the pieces must also be mistaken by professional musicians, composers, and conductors as one of Brahms' own works through a quantitative survey. Although there have been many examples of AI models generating music in the style of particular composers, no models have been created to generate the work of Brahms. The lack of a detailed computational analysis of Brahms shows a gap in the study of romantic period composers, which Brahms was a key figure of [1]. According to studies taken, computer-generated music has traditionally only sounded human-like when short excerpts were created and struggled with the complex motifs and harmonic cadences of romantic period piano music. This could be down to them being poor at handling higher-level musical structures due to the models only learning how to play the next note according to the previous [2]. In their paper, Child et al. developed a sparse transformer and stated that it was able to extract complex patterns from sequences up to 30 times longer than possible previously [3]. Likewise, Hawthrone et al. developed a Perceiver AR model which had the ability to effectively handle longer sequences with an improved memory efficiency [4]. After listening to the examples from the papers, the generated pieces still consisted of basic harmonic and rhythmic structures and struggled with the complex motifs and harmonic cadences of romantic period piano music. The importance of the research problem not only addresses AI's ability to generate music but also highlights the potential significance of how music could be composed in the future, particularly for those with no previous musical knowledge [5]. The research assumes that neural network models can already be trained to learn the general characteristics and patterns of various musical composers. To help with producing optimal results, the selected MIDI files for the dataset were exact replications of Brahms' piano music without errors and inconsistencies.

Based on the issues raised above, this study focused on understanding what configurations of the neural network models performed best when tasked with producing music in the style of Brahms. Therefore, the research question is:

To what extent can the accuracy of various Neural Network Models, trained with Long Short-Term Memory *and numerous Attention mechanisms, be significantly improved by augmenting MIDI files containing the compositional works of Johannes Brahms with an augmentation pipeline to generate pieces of music that are mistaken by professional musicians, composers and conductors as one of Brahms' own works?*

Due to conclusions from studies taken and the general state of knowledge at the time of beginning the project, the null hypothesis for this research project is;

H0: Neural network models cannot generate piano works to the same level of musicality and emotion as Brahms. Due to this, generated pieces will not be statistically similar through music information retrieval or mistaken as a work of Brahms by professional musicians, composers and conductors through a quantitative survey using Likert Scales.

Through the utilisation of an augmentation pipeline to expand the MIDI dataset containing the compositional piano works of Johannes Brahms, more musical variations could be created including transposition, rhythmical and note durations. In addition to this, various preprocessing techniques including track splitting, quantisation and normalisation could help make the MIDI files more readable for the models. This provides an alternate hypothesis;

H1: If an augmentation pipeline is utilised to expand a MIDI dataset of pre-processed files containing the piano works of Johannes Brahms, then various neural network models trained with Long Short-Term Memory and numerous Attention mechanisms could generate pieces of music that is statistically similar to Brahms and could be mistaken as one of Brahms' own piano works by professional musicians, composers and conductors through a quantitative survey using Likert Scales and various Independent-Samples T-Tests and Hotelling's T^2 Tests being implemented to determine whether the p-value is > 0.05 in order to reject the null hypothesis.

This paper contains a total of four sections. Section 2 describes the experiment design, methodology and how the dataset was prepared and pre-processed. Along with this, the neural network models obtained for the project and MIR functions will be explained as well as ethical considerations. Section 3 analyses and evaluates the results of the quantitative experiment and survey to determine if the experiments provide evidence that the null hypothesis is incorrect. Section 4 summarises what has been learnt and proposes recommendations and adjustments for future studies.

II. DESIGN AND METHODOLOGY

The research project was carried out through three stages. Firstly, data was collected, pre-processed, and augmented. The dataset contained 67 MIDI files of Johannes Brahms' piano works and was obtained for offline manipulation from *Classical Archives* and *MIDIworld*. An augmentation pipeline was employed to create variations in melodies, tempo, rhythm, and transpositions. This was followed by obtaining and training various neural network models with the augmented dataset. The models were analysed for training accuracy and loss to determine the best configurations. After training, the best generated examples from each model were analysed through various MIR functions using *MIDIToolbox* and *MIRToolbox* to determine the best performing models [6][7]. Finally, the best models generated pieces of music that were evaluated against the pieces of Brahms' repertoire for statistical equivalence. Along with this, the same generated pieces were used in a quantitative listener survey to test professional musicians, composers and conductors on whether they could differentiate between the generated pieces against the Brahms original. All quantitative experiments involved evaluation to test the research hypothesis through Independent-Samples T-tests and Hotelling's T^2 Tests.

A. Preparation and Preprocessing of Dataset

The preparation and preprocessing of the dataset involved numerous steps to optimise the potential of training the neural network models. Track splitting involved dividing the MIDI tracks into smaller segments of 30 second clips to make the tracks more digestible for the models in training [8]. The conversion of MIDI files into a single track allowed for further simplicity in the files. Normalisation was performed to ensure that the audio levels of the files were at the same amplitude to provide consistent values for training and evaluation tasks. Finally, all the MIDI files were quantized to semiquavers to adjust the timings of notes and align them with the correct timing to ensure consistency in rhythm [9]. An augmentation pipeline was utilised to create variations in melodies, tempo, rhythm and transpositions. Several techniques were used to increase the dataset size. Time-stretching was applied to make each MIDI file 5% faster or slower. Another method was to transpose each of the MIDI files so that the pitches would be raised or lowered by a third [8]. Doing these increased the dataset by 500% with a total of 445 tracks. In comparison with prior studies, data preparation and preprocessing was influenced by previous research, which was collected, adapted and integrated into this paper.

B. Neural Network Models

Numerous Neural Network Architectures were obtained for offline manipulation and trained with the augmented Brahms MIDI dataset. A Recurrent Neural Network was acquired from TensorFlow [9]. A RNN was described by IBM as "a type of artificial neural network which used sequential data or time series data" [10]. Sequential data was utilised to predict the next output based on previous elements in the sequence. RNNs suffer from a vanishing gradient problem. With the neural network using the gradient descent algorithm to update the weight, the gradients therefore decreased in growth the further down the layers the network progressed. A solution to this problem is the use of LSTM which utilises gating mechanisms to control the movement of information and gradients to allow for the network to learn and maintain information over longer sequences. [11] An LSTM-based RNN was obtained from Huang et al. [12]. Various Transformer models containing different attention mechanisms were obtained from Project Los Angeles [13]. Proposed by Google researchers Vaswani et al., transformer models do not rely on recurring processing of data. Instead, they operate on an attention mechanism [14]. Attention allows for neural networks to concentrate on particular parts of the input. The attention mechanisms tested were:

- Self-Attention – Processes inputs in the same sequence, enabling the model to capture dependencies within the input [15].
- Relative Attention – The relative position of tokens is considered based on the similarity of other tokens in the sequence [16].
- Local Windowed Attention – Restricts attention to a fixed window of tokens, enabling the model to focus on nearby information [17].
- Relative Self-Attention – Combination of Self and Relative attention allowing the model to focus on relevant information based on the positional relationships of tokens [14].
- Sparse Attention – Attends to a subset of tokens instead of the entire sequence to improve efficiency while retaining information [18].

A Perceiver AR model was also obtained from Project Los Angeles. Seen as an improvement to the Transformer model using latent array to distinguish the size of inputs and outputs, allowed for the model to efferently handle longer sequences with an improved memory efficiency [19].

C. *Quantitative Experiment & Survey*

A quantitative experiment and survey were conducted to test the neural network model's ability to replicate the musical characteristics and motifs of Brahms' piano works. Various MIR variables were utilised to gather various quantitative data from the models to test against not only each other but the original works of Brahms. These variables included:

- Entropy – The measure of uncertainty or unpredictability
- Duration Distribution – The statistical analysis of note durations as well as silence
- Pitch Class Distribution – The evaluation of frequency of musical pitches
- Mean Roughness – The measure of dissonance or clashing sounds
- Global Energy – The total amount of energy held within a waveform
- Normalised Pairwise Variability Index (nPVI) – The analysis of variability between successive durations
- Pulsation Clarity – The strength of the rhythmic pulse in a piece of music

The quantitative survey contained a total of 10 questions all containing the question *"Rate the likelihood that this piece was composed by Johannes Brahms as opposed to being generated by AI"* In random order, 5 pieces contained the works of Brahms and 5 were generated by the models. Answers consisted of a Likert Scale of 5 values ranging from *Definitely generated by AI* to *Definitely generated by Brahms.* Members of the Irish Defence Forces School of Music, RTÉ Concert Orchestra and National Symphony Orchestra were recruited for the survey to provide professional expertise in the subject. Participants were selected based on their extensive experience in classical music, including familiarity with Brahms' works, having performed his pieces in the past.

Using IBM's SPSS software, quantitative values from the MIR functions *MIRToolbox* and *MIDIToolbox* were tested to obtain p-values. Various Independent-Samples T-Tests and Hotelling's T^2 tests were implemented to evaluate musical variables to determine if there was statistical significance between Brahms' piano works and the generated pieces from the AI models.

D. *Ethical Considerations*

Several ethical considerations were adhered to in order to correctly conduct research including copyright issues and collection of data from survey participants A total of 67 pieces of music were obtained for offline manipulation for the MIDI dataset. According to German Federal Law Gazette, copyright protection for musical and artistic works expired 70 years after the death of the creator. With Brahms passing away in 1897, his compositions therefore resided in the public domain. The use of MIDI files also prevented any issues with performers rights as recordings of Brahms' works were not being used. All the neural network models obtained for testing were open-source and ran under the Apache 2.0 License. MIR tasks were performed using the *MATLAB* functions *MIRToolbox* and *MIDIToolbox*. To utilise the tools, *MATLAB* had to be downloaded free of charge under the GNU General Public License. Ethical considerations were vital when collecting data from personnel for the quantitative survey. Participation in the survey was voluntary and those who chose to partake were informed of the purpose of the study. No personal information was required from participants and the confidentiality of participants was guaranteed from the designer of the survey. The results of the survey were not tampered with and therefore were accurate.

III. EVALUATION

Overfitting issues were observed during training, with the LSTM and Perceiver AR models initially replicating the training material excessively instead of generating unique motifs. To mitigate this, the dataset was further augmented by transposing musical phrases and altering rhythms through quantisation to introduce additional variations in tempo and phrasing. Experimenting with different temperature values for each model allowed for control over the balance between simplicity and randomisation. With lower temperatures producing more basic outputs that resembled the training data, while higher temperatures encouraged greater diversity but could lead to compositions with less structure.

The model performance metrics stated that the transformer model with relative self-attention scored the best training loss and accuracy with scores of 0.015 and 0.995 respectively. The recurrent neural network scored worst in terms of training accuracy with a score of 0.628.

Through quantitative experimentation and a survey, the neural network models utilised for the project were evaluated extensively in a numerical form and through professional human judgement in order to confirm or refute the research hypothesis that neural network models could generate pieces of music with statistically similar musical characteristics and emotion as Brahms. To conduct a fair experiment, each model had to generate a two-minute-long piece which contained 300 prime tokens (30 seconds) from the beginning of six of Brahms' piano works. These generated pieces were then compared with the first two minutes of the original Brahms piece to evaluate the evolution of the generated pieces and determine their ability to maintain the style and structure of Brahms. MIR evaluation concluded that the Transformer models with self-attention, local windowed attention and relative global attention performed best in generating music most similar to the Brahms original.

Several statistical tests were conducted on the best performing models to obtain p-values to test the research hypothesis. The variables Entropy, nPVI, Global Energy, Mean Roughness and Pulsation Clarity were tested with an Independent-Samples t-test and the variables Duration Distribution and Pitch Class Distribution were tested with a Hotelling's T^2 test. Testing concluded that no statistical significance was found with most of the variables therefore supporting the alternative hypothesis that neural network models can produce music similar to Brahms. However, the variables Entropy and Global Energy were deemed to contain statistical significance within them stating that further work must be done to improve complexity, uncertainty and energy to a similar level to Brahms. Using *MIRToolbox*, the waveforms of the generated pieces were evaluated.

Figure 1 shows the brightness curve of Brahms' 2 Rhapsodies No. 1 alongside the generated piece from the transformer model with local windowed attention. The generated piece displays a greater variance of frequencies, resulting in a higher entropy score.

The difference in global energy is depicted in Figure 2 where the temporal evolution curve reveals a much greater variance in the generated piece compared to Brahms' original work. While Brahms' piece maintains a steady flow of increase and decrease of tension, the generated piece has much greater variations in timbre and harmonics throughout. The evaluation of these waveforms determined that the entropy and global energy values may have been much higher than Brahms' works due to the *MuseNet* inspired workflow the transformer models undertook causing the pieces to be generated in blocks and therefore sound unnatural.

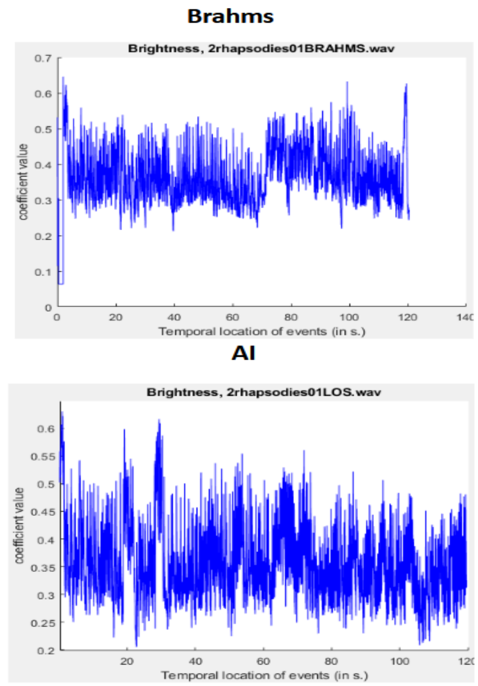


Figure 1. Brightness Curve between Brahms and AI

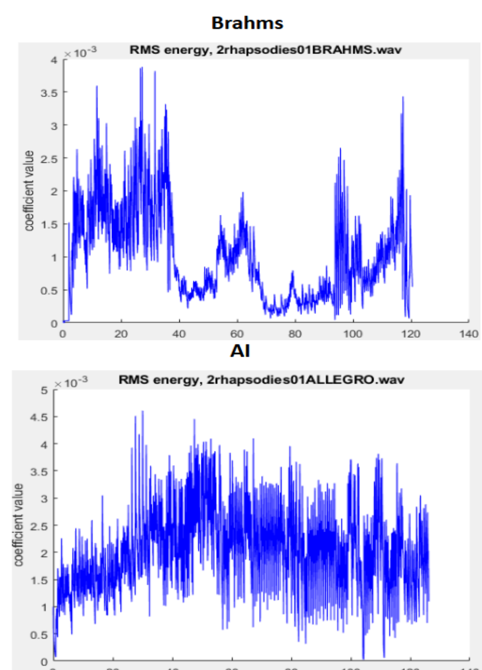


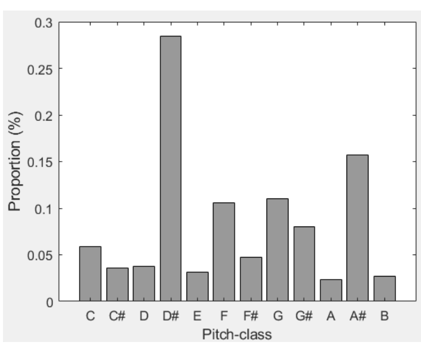
Figure 2. Temporal Evolution curves for Brahms and AI

Figure 3 displays the Pitch Class Distribution in the form of a box plot for Brahms' *3 Intermezzi No.1* and the generated piece from the transformer model with local windowed attention. With the piece being in the key of D# Major, which has an enharmonic equivalent of C minor, in order to stay within the key signature, the majority of the notes must be from the following triads:

- D# Major – D#, G, A#
- C Minor – C, D#, G

The pitch class distribution showed that both the original Brahms piece and AI generated were kept within these guidelines, with particular emphasis being placed on the notes D# and G as they feature in the triads of both D# major and C minor. Although the transformer model focused on the notes within the two triads to ensure consistency in the key signature, it was apparent that the model was reluctant to incorporate accidentals to further add colour to the piece. The use of extended chords was a key factor of the romantic period in which Brahms lived in and it was an era that bridged the gap between classical and modern music. *MIRToolbox* was able to identify both the Brahms and AI pieces to be in the key of D# major. With the function *mirmode,* it identified that the generated piece scored a higher probability of being in a major key than the Brahms original. Although it was positive that the generated piece was able to keep within the key signature for longer generated sequences, it does show an inability to evolve melodically into different harmonics.

Brahm



AI

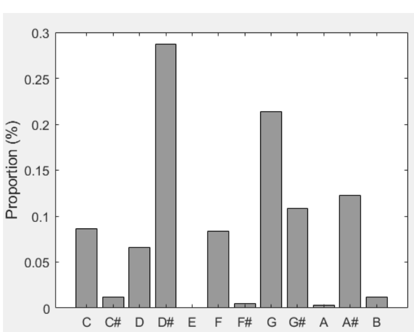


Fig. 3 Box Plot of the Pitch Class Distribution between Brahms and AI

A quantitative survey was also conducted to obtain human evaluation on 30 second clips of generated pieces from the models against the original works of Brahms. A total of 56 participants featuring only professional musicians, composers and conductors, displayed difficulty in recognising the distinction between the Brahms and generated pieces provided, with the majority incorrectly identifying two of the generated pieces as one of Brahms' own works.

Table 1 shows the percentages for each of the questions alongside whether the track was that of Brahms or AI. The total score was also calculated with the number of responses per answer being multiplied dependant on how similar to Brahms it was scored with *Definetly generated by AI* scoring 1 and *Definitely generated by Brahms* scoring 5. This was designed so that uncertainty was treated as a reward for the AI models as they still had not been identified as not Brahms.

An Independent-Samples T-Test was conducted which stated that there was no statistical significance between the Brahms and generated pieces therefore supporting the alternative hypothesis that neural network models trained with an augmented and pre-processed dataset could generate music at the level of musicality and emotion as of Brahms so much that through a quantitative survey the difference could not be identified by professional musicians, composers and conductors.

TABLE 1. QUANTITATIVE SURVEY RESULTS

	AI	Probably AI	Unsure	Probably Brahms	Brahms	Total Score
Brahms	3.64%	5.45%	3.64%	36.36%	**50.91%**	**237**
Brahms	12.50%	25.00%	3.57%	**37.50%**	21.43%	**185**
Brahms	0%	23.21%	7.14%	**39.29%**	30.36%	**211**
Brahms	17.86%	**44.64%**	3.57%	28.57%	5.36%	**145**
Brahms	7.27%	7.27%	18.18%	**41.82%**	25.45%	**207**
AI	23.21%	**32.14%**	12.50%	23.21%	8.93%	**147**
AI	10.71%	19.64%	19.64%	**28.57%**	21.43%	**185**
AI	5.36%	**41.07%**	14.29%	32.14%	7.14%	**165**
AI	14.55%	**34.55%**	5.45%	32.73%	12.73%	**165**
AI	7.27%	30.91%	7.27%	**36.36%**	18.18%	**183**

IV. CONCLUSION AND FUTURE WORK

The work conducted in this paper differed to previous studies as it trained various neural network models with a dataset containing the piano works of Brahms, an important figure of the Romantic Period in Classical Music. By doing this, a gap in the research was addressed by analysing a very important composer in an era of classical music were harmonic and rhythmic structures began to diverge from the traditional aspects of Renaissance and Classical Period music while also bridging the gap between traditionalism and modernism. While the literature review stated a gap in the research that previous models struggled with complex motifs and harmonic cadences of romantic period piano music. This paper found that statistical testing in various

 ISBN: 978-1-68558-262-3

musical categories stated there was not statistical significance between the Brahms and AI generated pieces. A quantitative survey containing participants who were educated in the subject mistook two of the neural network models generated pieces as Brahms' own works suggesting the model's ability to generate pieces of music with the complexity in rhythmic and harmonic characteristics of Brahms.

The results from the research carried out suggest that transformer models with self-attention, relative self-attention and local windowed attention were able to generate various characteristics to a statistically similar level to Brahms with a dataset of his piano works by utilising an augmentation pipeline and various preprocessing techniques. A couple of musical characteristics however proved to be statistically significant to Brahms, these being entropy and global energy. This concludes that while the transformer models are able to replicate a vast amount of Brahms' compositional traits, they still fall behind in reproducing the rhythmical and harmonical complexities, uncertainties and global energy levels of Brahms' works. The possibility of increasing the dataset to the orchestral, ensemble and choral works of Brahms could greatly increase the abilities of generated music from just solo piano works. This also could adhere to limitations regarding a small dataset and therefore improve accuracies in entropy and global energy from the generated pieces.

While participants noted difficulty in distinguishing between the Brahms and AI pieces, some commented that the use of MIDI files made all the music sound robotic and therefore made it even more challenging to differentiate between the two. Future work could focus on converting the generated pieces into musical notation and having a professional pianist perform them. This would enable an experiment to assess the generated music on an acoustic piano, the instrument it was originally intended for.

REFERENCES

[1] J. D. Fernández and F. Vico, "AI Methods in Algorithmic Composition: A Comprehensive Survey," *Journal of Artificial Intelligence Research*, no. 48, pp. 513-582, 2013. https://doi.org/10.48550/arXiv.1402.0585

[2] K. Zheng, R. Meng, C. Zheng, X. Li, J. Sang, J. Cai. and J. Wang, "EmotionBox: a music-element-driven emotional music generation system using Recurrent Neural Network," *Frontiers in Psychology*, no. 13, pp. 1-14, 2021. https://doi.org/10.3389/fpsyg.2022.841926

[3] R. Child, S. Gray, A. Radford and I. Sutskever, "Generating Long Sequences with Sparse Transformers. *PsyArXiv,* no. 1, pp. 1-10, 2019. https://doi.org/10.48550/arXiv.1904.10509

[4] C. Hawthorne, A. Jaegle, C. Cangea, S. Borgeaud, C. Nash, M. Malinowski, S. Dieleman, O. Vinyals, M. Botvinick, I. Simon, H. Sheahan, N. Zeghidour, J. B. Alayrac, J. Carreira, and J. Engel, "General-purpose, long-context autoregressive modelling with Perceiver AR," *In Proc of the International Conference on Machine Learning*, no. 39, pp. 1-24, 2022. https://doi.org/10.48550/arXiv.2202.07765

[5] E. Deruty, M. Grachten, S. Lattner, J. Nistal and C. Aouameur, "On the Development and Practice of AI technology for Contemporary Popular Music Production," *Transactions of the International Society for Music Information Retrieval,* no. 5, pp. 35-49, 2022. https://doi.org/10.5334/tismir.100

[6] O. Lartillot, P. Toiviainen, P. Saari and T. Eerola, "MIRtoolbox," 2007. [Online]. Available: https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox

[7] T. Eerola and P. Toiviainen, "MIDI Toolbox: MATLAB Tools for Music Research," 2004. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6e06906ca1ba0bf0ac8f2cb1a929f3be95eeadfa

[8] S. Oore, I. Simon, S. Dieleman, D. Eck and K. Simonyan, "This Time With Feeling: Learning Expressive Musical Performance," *Neural Computing and Applications,* no. 32, pp. 995-967, 2018. https://doi.org/10.1007/s00521-018-3758-9

[9] J. A. Franklin, "Jazz Melody Generation from Recurrent Network Learning of Several Human Melodies," *In Proc of the International Florida Artificial Intelligence Research Society Conference,* no. 18, pp.57-62, 2005. https://cdn.aaai.org/FLAIRS/2005/Flairs05-010.pdf

[10] TensorFlow, "Generate music with an RNN," n.d. [Online]. Available: https://www.tensorflow.org/tutorials/audio/music_generation

[11] IBM, "What are recurrent neural networks?" n.d. [Online] Available: https://ibm.com/topics/recurrent-neural-networks

[12] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation,* no. 9, pp. 1735-1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735J.

[13] C. Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu and D. Eck, "Music Transformer: Generating Music with Long-Term Structure," *Proceedings of the International Conference on Learning Representations,* no. 17, pp. 1-14, 2018. https://doi.org/10.48550/arXiv.1809.04281

[14] A. Sigalov, "Project Los Angeles" 2019. [Online]. Available: https://github.com/asigalov61

[15] A. K. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention is All you Need," *Advances in Neural Information Processing Systems,* no. 31, pp. 1-15, 2017. https://doi.org/10.48550/arXiv.1706.03762

[16] A. K. Huang, P.A. Szerlip, M. E. Norton, T.A. Brindle, Z. Merritt and K.O. Stanley, "Visualizing music self-attention" *In Proc of the Conference on Neural Information Processing Systems,* no.32, pp. 1-5, 2018. https://openreview.net/pdf?id=ryfxVNEajm

[17] P. Shaw, J. Uszkoreit and A. Vaswani, "Self-Attention with Relative Positional Representations," *In Proc of the 2018 Conference of the North American Chaper of the Association for Computational Linguistics: Human Language Technologies,* no.2, pp. 464-468, 2018. https://doi.org/10.18653/v1/N18-2074

[18] Z. Liu, Y. Lin, H. Hu, Y. Wei, Z. Zhang, S. Lin and B. Guo, "Swim Transformer: Hierarchical Vision Transformer using Shifted Windows," *In Proc of the IEEE/CVF International Conference on Computer Vision,* no.23, pp. 9992-10002, 2021. https://doi.org/10.48550/arXiv.2103.14030

[19] A. Jaegle, F. Gimeno, A. Brock, A. Zisserman, O. Vinyals and J. Carreira, "Perceiver: General Perception with Iterative Attention," *In Proc of the International Conference on Machine Learning,* no. 38, pp. 1-43, 2021. https://doi.org/10.48550/arXiv.2103.03206