Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/68/20241408

Analysis the approaches and applications for jazz music composing based on machine learning

Ruoxi Du

College of humanities, Xidian University, Xi'an, 710071, China

rxdu@stu.xidian.edu.cn

Abstract. As a matter of fact, computer composing is a hot topic for study in recent years. To be specific, Jazz, one of the essential music genres, has a complex and irregular musical structure. Current researchers are focusing on how to use models to generate expressive and innovative jazz. This study first summarizes in detail the characteristics of the musical structure of jazz and the unique structures of jazz music that are the difficulties of model training. At the same time, it then analyzes the feasibility and application of several popular machine learning models in jazz music composition. Finally, this study combines the current development of the field of computer composition with the challenges faced by the field of computer-generated jazz, as well as the direction of the next step in the development of continued efforts to explore in-depth, in the hope that this will provide researchers with new research perspectives, and to promote the development of new forms of jazz music can flourish in the age of intelligent machines.

Keywords: Machine learning, music composing, Jazz.

1. Introduction

Computer music is not new in recent years. As early as the 18th century, Turing discovered that when a computer makes a repetitive "tick" sound at a specific frequency, it produces different notes. Soon after, Programmer Strachey, whom Turing influenced, successfully wrote the first British computer music, which laid the foundation for the subsequent development of computer music [1]. In the 20th century, with computer technology's continuous development and maturity, completely new models and algorithms were born. Since the beginning of this century, more and more researchers have begun to explore more possibilities of music in the new era. For example, shortly after the theory of backpropagation was proposed, it had an impact on the development of many models [2]. Researchers soon began to explore the application of RNN in various fields and music. For instance, by simulating non-musician listeners' perception of tonal structure in the study, the researchers found that neural networks can effectively learn the structure and rules of music [3]. Shortly after 1997, the birth of Long Short-Term Memory (LSTM) solved vanishing gradient issue and exploding gradient problem in traditional recurrent neural networks [4], and this technological breakthrough also gave computer music a better music generation effect.

Advances in deep learning have significantly improved the evolution of modern computer music, especially in areas such as music generation and style transformation. Nevertheless, traditional machine learning models have yet to be phased out by the rise of deep learning; instead, they continue to play a crucial role in areas such as music analysis and modeling. For example, Verbeurgt et al., in their 2004

DOI: 10.54254/2755-2721/68/20241408

study, used Markov chains to analyze music data from three latitudes (time, pitch, and duration) to learn and extract more complex patterns from classical music sequences [5], which enhanced the smoothing of musical transitions. At the same time, it promotes the further development of the algorithmic composition model. Similarly, in 2001, Emanuele and Simoncelli used Hidden Markov Models (HMM) to collect and train datasets of different musical styles, and they comparatively analyzed the differences between humans and computers in music recognition while they looked ahead to the future megatrends for the development of more accurate models of memory [6]. These studies confirm the effectiveness of traditional machine learning models in capturing and analyzing musical structure, as well as their continued unique value and promise for application in the current deep learning-dominated research environment.

In this context, jazz, as a form of music that people have loved for centuries, has become a challenge in the field of computer music research because of its unique musical structure. Musical tension and emotional expression are central elements when composing jazz, and at the same time, this has attracted the attention of researchers. For example, in 2005 Ramirez, R et al. summarized the inductive logic of jazz, leading to a further in-depth exploration of how expressiveness in jazz can be represented by computers [7]. Subsequently, Giraldo and Ramirez employed the Ripper algorithm to categorize the collected data while combining multiple machine learning models to allow for more emotional expression in computer-generated jazz [8, 9]. The improvisation characteristic of jazz is also an essential part of jazz because of its special irregular melody trend. When the computer is used to accompany musicians in real-time, the model needs to accurately match the correct harmony in real-time. Deep neural network modeling is a suitable method [10]. These innovative studies on jazz not only enrich the computer music field theory but also create a solid basis for the subsequent development of jazz.

The writing of this article was inspired by several sources. First, although there has been much research on machine learning in different musical directions over the past decade, there is currently less comprehensive research focused on jazz. Therefore, this paper seeks to close this disparity and offer a comprehensive perspective for academics and developers by systematically analyzing important research and advances in the field. Second, given that artificial intelligence is maturing in terms of its application in music composition, this paper also aims to explore the current prospects for the practical application of machine learning in jazz composition. In the remaining Sec. 2, this paper describes the characteristics of jazz, Sec. 3 is the research of the current popular models and the practical application analysis in jazz, Sec. 4 is to analyze the current limitations and future development, and Sec. 5 is the conclusion.

2. Characteristics of Jazz

Historically, jazz has been influenced by African and European music, so many other musical styles have been included in jazz, such as the blues, the functional harmony of the 18th century, the chromatic harmony of the late romantic attention of the 19th century. These diverse and complex harmonic systems converge into a short dozen measure in jazz, which allows jazz to grasp the listener's auditory nerve in a very short time. Swing is an important element in jazz [11], and it comes from four main areas: rhythm, emphasis, harmony, and improvisation. The rhythmic distribution of traditional music usually adopts an evenly divided structure, which emphasizes the precision and regularity of the arrangement of the movement, so that it will reflect the aesthetics of classical music in terms of structure and formal presentation. For instance, in Bach's and Mozart's works, there is a strict rhythmic treatment that makes the musical direction very clear and predictable. However, in jazz triplets (e.g., seen from Fig. 1), the first two eighth notes are played together, which means that the first t two notes will account for 2/3 of the total time value. In this way, the rhythm ratio will roughly change from 1:1 to 2:1, creating an asymmetric rhythm pattern, giving the audience the feeling of rocking back and forth, ups and downs.

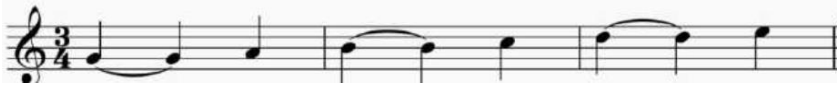

Figure 1. Triplet example (Photo/Picture credit: Original).

DOI: 10.54254/2755-2721/68/20241408

Jazz, because of its blend of blues, will have some similarities to blue in terms of musical arrangements, like swing and shuffle. Shuffle is very similar to Swing in that they both have a musical structure that is long before and short after, which can occasionally make it challenging to distinguish one from the other. Yet people can distinguish between them by the jazz emphasis; the jazz emphasis is in the middle of a triplet of strong beats, which gives the jazz more lightness than the blue. The unique harmonic structure is also a very attractive part of jazz. Extended chords such as seventh, ninth, and eleventh chords are widely used in jazz. These chords are more complex in terms of chord composition than the more common triads in classical music (major and minor). At the same time, the use of dissonance in jazz is also more frequent, and all these harmonic compositions make the music of jazz more colorful, to achieve with the creator's intention to use the music to reveal the context of the current era.

Finally, improvisation is the soul of jazz. Despite there have been a number of improvisations in European music history, jazz improvisation is very different from them in that it incorporates many elements of African music, like call and response, in which each player responds instantly to the tune they hear, without repeating the previous section. Jazz players deliberately learn how to construct their own jazz phrases in the early stages of learning, ensuring that they will think in jazz language when improvising [12], and computer machine learning is similar to them to a certain extent. However, when machine learning models are used to create melodies or follow real-time accompaniments, these complex harmonic structures and strongly uncertain trends of improvisational music will be greatly influenced. All of them have certain requirements for model training, which is an important technical difficulty for algorithm developers.

3. Model and application

This section will sort out the internal core mechanisms of traditional machine learning and its subset of deep learning models widely applied in the field of jazz composition. It will also analyze the current applications of jazz generation in conjunction with the models.

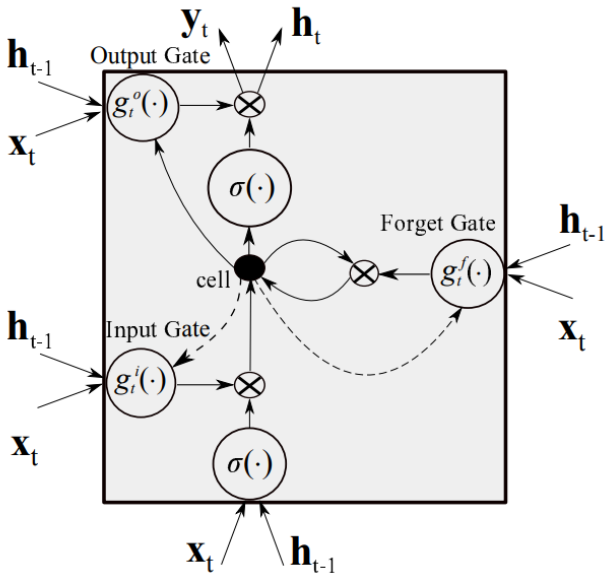

Figure 2. Schematic of an LSTM Cell Architecture [13].

3.1. LSTM

By putting in place gate controls, namely, the forgetting gate, input gate, and output gate. LSTM networks can control forgotten historical information, manage input data, and adjust output results, respectively, as shown in the Fig. 2 [13]. Meanwhile, LSTM also introduces a memory cell for storing

DOI: 10.54254/2755-2721/68/20241408

historical information and controlling the information flow, ensuring information stability in the transmission process. All these functions further significantly enhance the model's capacity to deal with long-term dependent information, thus solving the issues of vanishing gradients and exploding gradients in traditional RNN [4]. In jazz generation, due to its dynamically changing nature of improvisation and complex musical structures, training music generation models requires using long time series information to capture these structures accurately. Compared with other neural network structures (RNN, CNN, DBN, etc.), LSTM is well suited for jazz generation due to its significant advantages in learning long-term dependencies. It can effectively restore the continuity and complexity of jazz, which makes the generated music maintain the style of jazz as well as enrich the expression of the music [14].

3.2. CNN

CNN were first effectively implemented in a handwritten character recognition system for the United States Post Office by LeCun, who created them in 1989 [15]. The several modules in a CNN are very important for data extraction. Illustrated in Fig. 3, the first Convolutional Layer (C1) generates 6 feature mappings of 28x28 each, and then the subsequent Convolutional Layer (C3) further processes and generates more complex advanced feature mappings. Responsibility for diminishing the spatial dimensions of feature maps falls to the pooling layer, which consequently decreases computational demands and achieves some level of translation invariance. Also, the Pooling Layer will always follow each Convolutional Layer, as shown in the Fig. 3 (S2 and S4) [16]. Finally, the Fully Connected Layer synthesizes the local features for final classification or other task decisions, as illustrated in the Fig. 3, the two Fully Connected Layer (C5 and F6) ended up outputting ten classification results, one for each category. In addition to these three important modules, the Activation Function, the Normalization Layer, and the Dropout Layer also play important roles in CNN.

In specific music applications, although CNN can also be used to generate music, its more prominent advantage is in areas such as music genre classification. Because of its powerful feature extraction and hierarchical learning capabilities, it makes it possible to parse and categorize a variety of complex musical structures [16], such as the complex structures of jazz. In jazz, where a complete composition usually includes a variety of instruments and improvisations, CNN can effectively extract useful features from complex jazz audio by using Multiple Sequential Convolutional Layers and a Max Pooling Layer. These Convolutional Layers can be used to identify jazz-localized audio features such as pitch, rhythm, and timbre, while the Max Pooling Layer helps to diminish the feature dimensionality and enhances the model's generalization capability. This feature allows CNN to have a certain degree of usability in applications that deal with the specific structure of jazz, jazz analysis, or jazz generation in combination with other algorithms (LSTM, GAN, etc.) [17].

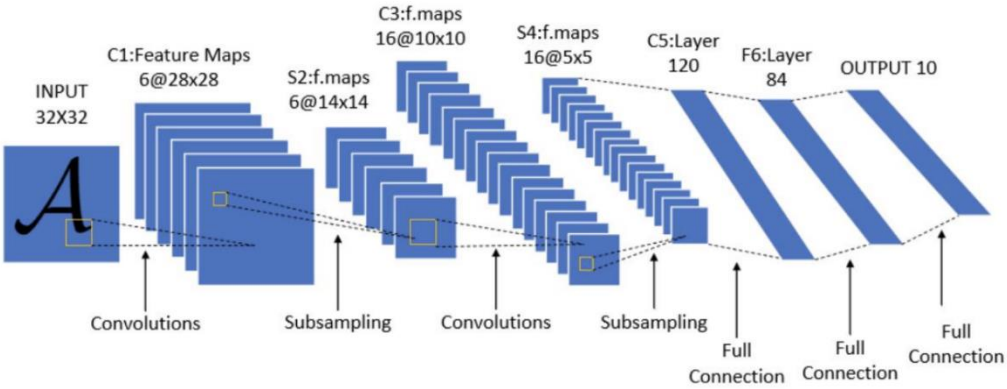

Figure 3. CNN Architecture Diagram [16].

3.3. GAN

GAN, developed by Ian Goodfellow in 2014, is a probabilistic generative model based on game theory, with "Generators" and "Discriminators" as its core components. In the training process of GAN, take

DOI: 10.54254/2755-2721/68/20241408

image generation as an example, these two networks will learn by confronting each other, The Discriminator is trying to figure out how to tell the difference between the real data and the fake data that the Generator is producing, and the Generator's objective is to create as realistic images as possible to fool it, this confrontation process is constantly cycling, as a way to improve the performance of the two networks and to produce higher-quality data samples, as in the process shown in the Fig. 4 [18].

In the field of computer composition, having GAN trained in combination with other models is very effective in generating high-quality musical compositions. For instance, Modrzejewski et al, use CNN as Generators and Discriminators in GAN. They took CNN to efficiently extract spatial and hierarchical features from images, and use MIDI audio-converted images as data to train GAN models, an innovation that allows GAN to avoid repetitive phrases when generating jazz clips while still conveying the mood and ideas of the music in jazz [19]. This model innovation based on GAN provides new possibilities for the generative music field.

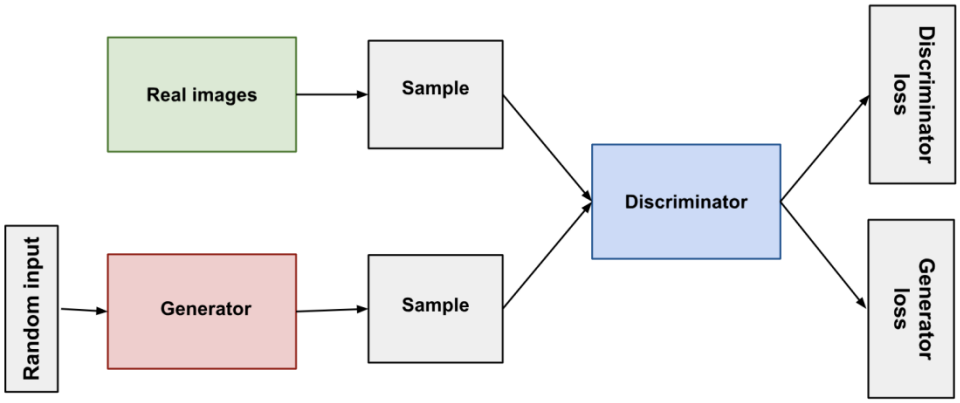

Figure 4. GAN Framework [18].

3.4. Transformer and Transformer-XL

In 2017 Google team proposed the Transformer model [20], which is a redesigned model for the weaknesses of RNN, solving the problems of low efficiency of RNN and defects in information transfer, it can be more efficient in dealing with sequence data to be able to learn the complex relationships between sequences. Transformer employs an Encoder-Decoder structure, as illustrated in the Fig. 5 for the left and right sections. The Encoder unit helps to stabilize the training process, accelerate convergence, and mitigate the vanishing gradient problem. The Decoder unit builds upon the Encoder by adding an additional multi-head masked attention component, which prevents the use of words that have not yet been predicted during forecasting, allowing the generation of the next word based solely on what is already known. Transformer's self-attention mechanism is very effective for addressing long-range dependencies and complex structures within musical compositions, and this mechanism allows the model to consider the complete input sequence information during the generation process as a way to create music with rich layers and coherent structures [21].

Transformer-XL is a variant of Transformer that adds a Segmentation Loop Mechanism and Relative Position Encoding to the original, and the addition of these core components allows the model to perform even better when dealing with tasks that rely on long-term memory, while Transformer-XL further solves some of the long temporal processing problems that still exist in Transformer. Transformer-XL also further addresses some of the long temporal processing issues that still exist in Transformer. The addition of these two core components gives Transformer-XL a great advantage in generating complex jazz. The segmentation mechanism significantly improves the model's ability to handle long sequences of data, allowing the model to accurately grasp musical changes and thematic development when generating long jazz compositions, and the context-aware ability of relative positional encoding allows the model to more accurately capture the structural relationships and chordal changes in the music.

DOI: 10.54254/2755-2721/68/20241408

structural relationships and chord changes in the music, improving the expressiveness and accuracy of the generated music [22].

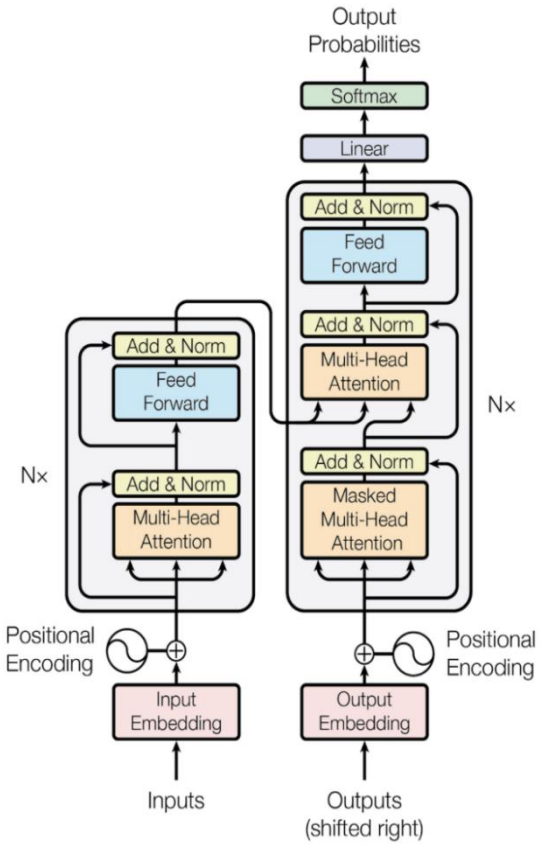

Figure 5. Transformer Model Architecture [21].

3.5. HMM

HMM is an extended model based on Markov Chain, which inherits the core "memorylessness" of Markov Chain while adding mechanisms that can handle unobservable (hidden) states. Transition Probability and Emission Probability are an important part of HMM, the former can reflect the internal logic and dynamics between the hidden states, how one state can be transformed to another state based on the internal rules of the model. And the latter associates the hidden states with the observed data, allowing one to further infer these states from the observed data [23]. The hidden state mechanism of HHM is well suited for learning the structure of musical compositions, and the following conclusions can be drawn by analyzing the Transition Probability formula as following:

$$A = [a_{ij}]_{N\times N}, a_{ij} = P(i_{t+1} = q_j | i_t = q_i) \quad (1)$$

$$B = [b_{ik}]_{N\times M}, b_{ik} = P(o_t = v_k | i_t = q_i) \quad (2)$$

Here, the element a_{ij} in Transition Matrix A indicates the likelihood of transitioning from the present state q_i to a succeeding stat q_j. In specific applications, by adjusting the probability values in Matrix A, it is possible to simulate the complex rhythmic and harmonic variability in jazz improvisation, providing a rich diversity for the hidden state sequence. For Eq. (2), the element b_{ik} in the emission probability matrix B denotes the probability of generating observation v_k given the state q_i. When generating jazz music, the emission probabilities influence the style and emotional content of Jazz, allowing the produced pieces to more closely resemble real jazz compositions.

DOI: 10.54254/2755-2721/68/20241408

According to the model analysis in the previous subsection, it can be seen that the current deep learning models show some potential in the field of music generation, but their dependence on a substantial volume of training data to achieve the effect of music generation is not the most suitable method for jazz generation, because of the irregularity of the characteristics of the jazz improvisation solo, so that the deep learning models cannot achieve the expected results, and the HMM in the limited size of the available dataset can be very good to complete the generation of the basis. The HMM can be used to predict the next possible note or chord and make a decision through the Viterbi algorithm, which can be very good for generating music that fits the jazz continuum [24].

4. Limitations and Prospects

Nowadays, with the rapid development of computer-generated music, many sophisticated models have been successfully Adopted to the creation of various music genres. In classical music, the DeepBach model developed by Hadjeres et al. using a dependency network approach with a non-sequential sampling strategy, and the BachBot model developed by Liang using a 3-layer stacked LSTM for sequence modeling, have been successful in generating very specialized Bach-style chorales [25, 26]. In the field of narrative genres of music like film, television and games, the music generation system called MorpheuS developed by Herremans et al. solves the long-term structural problem in automatic music generation and is well suited for generating polyphonic music with emotional and narrative qualities [27]. In pop music, popular piano music with emotional expression and melodic structure was successfully generated using a very advanced neural sequence model (Transformer-XL) by Huang, Yu-Siang et al. [28]. These cases above are just a small selection of classic examples from the many different musical genres that have been studied. Although there have been some breakthrough theories on jazz in recent years, such as Kaliakatsos-Papakostas et al.'s suggestion in 2023 that generating Jazz using a traditional machine learning model, HMM, would be more effective than a deep learning model [24], after comparing the results of the research on other musical genres in general, the current research on jazz is still lacking to a certain extent in relation to other styles.

Furthermore, the technical development difficulties that were once caused by the nature of jazz improvisation, where the harmonic structure is constantly changing within the piece [29], are no longer a major problem for researchers in today's technological boom, as many newly bred models like LSTM networks, Transformer models, CNN, and so on, can be a good solution to this problem. Web platforms developed specifically for generating jazz have also emerged, for instance, Ji-Sung Kim's Deepjazz built with two learning libraries for deep learning (Keras and Theano) [30]. However, compare to Chat GPT, a chatbot, MidJourney, an image generation platform, AIVA, an AI platform for customized soundtracks, etc., such large-scale commercialized platforms that support customized user interaction. The contemporary development of platforms for computer-generated jazz is still relatively rare in the current mainstream AI market.

To summarize, the future development of jazz should further strengthen the in-depth research in this field, and through the continuous development of advanced technology and algorithms, continue to explore how to maintain the structural framework of jazz music on the basis of the creation of more innovative and jazz connotation of the works. Simultaneously, researchers should concentrate on advancing user-oriented interactive jazz platforms. allowing users to create jazz music more conveniently, so that more young people can learn about jazz, and promote jazz to continue to become the mainstream in the contemporary era. In order to achieve this, the development of interactive platforms needs to focus on the intuitive operation of the user interface, as well as on the development of more customized jazz improvisation models like BebopNet [31], which is specifically dedicated to generating personalized and customized jazz improvisation models that match the tastes of the users in order to satisfy the creative needs of the different users when they are used by future clients. With these changes, the combination of jazz and artificial intelligence will allow jazz to continue to flourish in the 21st century in a new form of music and contribute to the development of the field of computer music.

DOI: 10.54254/2755-2721/68/20241408

5. Conclusion

To sum up, this study explains in detail the characteristics of harmony, rhythm, improvisation and other musical components in jazz, and simultaneously combines these musical characteristics for exploring the generative effects and applications of several types of models that are currently widely used in jazz. Through the analysis of these studies, it is found that although these models have certain capabilities in generating works that conform to the characteristics of jazz, compared to the amount of research on other music genres, the research on jazz is still apparently insufficient, and there is still a challenge in developing models that can generate jazz works with more deeper meanings and innovations of jazz. In addition, the construction of a jazz-specific interactive music generation platform is also a pressing issue at present. The birth of a commercialized jazz generation platform will promote the development of jazz in the new era and may become a major trend in future research. Jazz is one of the most important forms of music that has been influencing people for over a hundred years, and at the meantime, artificial intelligence composition is also an unstoppable development trend nowadays. This paper hope to provide future researchers with some theoretical assistance by combing the current modeling techniques and applications of previous research, so as to promote the continuation of jazz in a way that combines science and technology.

References

[1] Copeland B J and Long J 2017 Philosophical Explorations of the Legacy of Alan Turing: Turing vol 100 pp 189-218.
[2] Rumelhart D E, Hinton G E and Williams R J 1986 Nature vol 323 pp 533–536.
[3] Bharucha J J and Todd P M 1989 Computer Music Journal vol 13(4) pp 44–53.
[4] Hochreiter S and Schmidhuber J 1997 Neural Computation vol 9(8) pp 1735–1780.
[5] Verbeurgt K, Dinolfo M and Fayer M 2004 Innovations in Applied Artificial Intelligence: 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE 2004 Proceedings vol 17.
[6] Pollastri E and Simoncelli G 2001 Proceedings First International Conference on WEB Delivering of Music, WEDELMUSIC vol 2001.
[7] Ramirez R, et al. 2005 Discovering expressive transformation rules from saxophone jazz performances. Journal of New Music Research 34.4: 319–330.
[8] Giraldo S and Ramorez R 2018 Machine Learning and Music Generation pp 21–40.
[9] Giraldo S and Ramirez R 2016Frontiers in Psychology vol 7 p 198200.
[10] Kritsis K, Kylafi T, Kaliakatsos-Papakostas M, et al. 2021 Frontiers in Artificial Intelligence vol 3 p 508727.
[11] Gridley M, Maxham R and Hoff R 1989 Three approaches to defining jazz. The Musical Quarterly vol 73(4) pp 513–531.
[12] Berliner P F 2009 Thinking in jazz: The infinite art of improvisation. University of Chicago Press pp 144–161.
[13] Salehinejad H, Sankar S, Barfett J, et al. 2017 Recent advances in recurrent neural networks. arXiv preprint arXiv:1801.01078.
[14] Wang J, Wang X and Cai J 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) vol 1 pp 115-120.
[15] LeCun Y, Bottou L, Bengio Y, et al. 1998 Proceedings of the IEEE vol 86(11) pp 2278–2324.
[16] Choi K, Fazekas G, Sandler M, et al. 2017 IEEE International conference on acoustics, speech and signal processing (ICASSP) pp 2392-2396.
[17] Abeßer J, Chauhan J, Pillai P P, et al. 2021 29th European Signal Processing Conference (EUSIPCO) pp 361-365.
[18] Goodfellow I, Pouget-Abadie J, Mirza M, et al. 2014 Advances in neural information processing systems vol 27.
[19] Modrzejewski M, Dorobek M and Rokita P 2019 Artificial Intelligence and Soft Computing: 18th International Conference, ICAISC Proceedings Part I vol 18.

DOI: 10.54254/2755-2721/68/20241408

[20] Vaswani A, Shazeer N, Parmar N, et al. 2017 Advances in neural information processing systems p 30.
[21] Dong H W, Chen K, Dubnov S, et al. 2023 Multitrack music transformer. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp 1-5.
[22] Wu S L and Yang Y H 2020 The Jazz Transformer on the front line: Exploring the shortcomings of AI-composed music through quantitative measures. arXiv preprint arXiv:2008.01307.
[23] Rabiner L and Juang B 1986 An introduction to hidden Markov models. IEEE ASSP Magazine vol 3(1) pp 4-16.
[24] Kaliakatsos-Papakostas M, Velenis K, Pasias L, et al. 2023 Applied Sciences vol 13(3) 1338.
[25] Hadjeres G, Pachet F and Nielsen F 2017 International Conference on Machine Learning PMLR p 17.
[26] Liang F 2016 Bachbot: Automatic composition in the style of bach chorales. University of Cambridge vol 8 pp 19-48.
[27] Herremans D and Chew E 2017 IEEE Transactions on Affective Computing vol 10(4) pp 510-523.
[28] Huang Y S and Yang Y H 2020 Proceedings of the 28th ACM International Conference on Multimedia p 13.
[29] Giomi F and Ligabue M 1991 Journal of New Music Research vol 20.1 pp 47-64.
[30] Yadav PS, et al. 2022 A lightweight deep learning-based approach for Jazz music generation in MIDI format. Computational Intelligence and Neuroscience 2022.
[31] Hakimi S H, Bhonker N and El-Yaniv R2020 BebopNet: Deep Neural Models for Personalized Jazz Improvisations. ISMIR vol 2020.