International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

MUSIC: Metrics-based Understanding of Soundscapes in AI Composition

Hrishikesh Yadav[1]*, Prerak Joshi[2], Jay Oza[3]

Submitted:11/03/2024 **Revised**: 26/04/2024 **Accepted**: 03/05/2024

Abstract: The present research explores into a multidimensional examination of AI-generated piano music, using various techniques that incorporates numerous musical measures. Pitch contour analysis, static velocity assessment, harmonic analysis, rhythmic correctness quantification with a threshold value, harmonic compatibility, liveliness evaluation, amplitude envelope evaluation, RMS Energy, and zero-crossing analysis are all part of the suggested technique. This investigation examines how these variables together help to understand the performance and divergence of AI-generated music compared to original works, using MIDI files as the foundation for AI-generated compositions. The results are more helpful to identify whether the music is AI generated or human. We give an original perspective on evaluating AI-generated music that exceeds existing methodologies, revealing insight on the expanding landscape of artificial creativity in music composition by using this complete methodology.

Keywords: *musical measures, musical instrument digital interface, pitch contour analysis, static velocity assessment, harmonic analysis, rhythmic correctness quantification, threshold value, harmonic compatibility, amplitude envelope evaluation, RMS Energy, zero-crossing analysis, music composition.*

1. Introduction

The rise of AI-generated music has been made possible by the use of deep learning techniques such as Transformers, Generative Adversarial Network (GANs), Variational Autoencoders (VAEs) and Generative AI. These algorithms may create music by learning from vast datasets of existing music and then composing new compositions based on that information. The Transformer model, in particular, has grown in favour because of its capacity to weigh the relevance of various portions of the input and produce unified and high-quality music. The employment of these models has enormous implications for the music business, as they are utilised by composers and producers to make music and as a tool for their job. There are an expected more original and creative compositions in the future as AI-generated music develops further.

Through Meta releasing an open-source alternative called MusicGen and Google just making their text-to-music AI, MusicLM, the area of AI music is advancing quickly [1][2]. Both MusicGen and MusicLM are generative AI models that create short music clips based on text input they both use machine learning. Utilising userprovided descriptions such as "90s rock songs with electric guitar" or "180 bpm gabber tracks with microtonal syntqh leads," MusicGen may produce 15-second audio pieces [3]. A more sophisticated version of MusicGen is capable of creating clips up to 120 seconds long and recreating specific melodies from reference audio files. MusicGen will work hard to realise your musical ambitions, whether you want a thrash metal cover of My Heart Will Go On or an EDM remix.

Music evaluation metrics are tools for determining the quality of music based on a variety of parameters. These measures are crucial because they allow us to more objectively evaluate and enjoy music. Music is a type of art that is appreciated in many ways by people all over the world. It is a cultural activity that takes the shape of organised music and is enjoyed by everybody. However, not everyone appreciates music for what it is, rather than just the sound. Music assessment metrics are required to assist individuals in better understanding and evaluating music based on its fundamental parts of sound, melody, harmony, rhythm, structure or form, expression, and texture.

Music becoming a rich and multidimensional art form, sometimes resists simple categorization or evaluation. It's a highly subjective sensation that differs widely amongst people [4]. Because of this subjectivity, objectively evaluating the quality of music is difficult. Music evaluation measures are critical in tackling this issue. Metrics enable us to evaluate music in a more consistent and objective manner by measuring these characteristics. This is especially useful for artists, music producers, reviewers, and listeners who want to comprehend and enjoy music in a more sophisticated and methodical manner.

Music evaluation metrics are a helpful teaching tool that may help people gain a better knowledge of music [5]. These metrics deconstruct music into its constituent aspects,

[1] Thakur College of Engineering and Technology, Mumbai, Maharashtra, India
ORCID ID : 0009-0009-9584-6714
[2] Thakur College of Engineering and Technology, Mumbai, Maharashtra, India
ORCID ID : 0009-0005-5248-4953
[3] K J Somaiya Institute of Technology, Mumbai, Maharashtra, India
ORCID ID : 0009-0002-0102-7478

enabling granular examination of sound, melody, harmony, rhythm, structure or form, expressiveness, and texture. Individuals may better understand the complicated interplay that distinguishes each musical composition as they become more familiar with these aspects and their significance. This awareness not only improves one's capacity to listen critically, but it also develops a broader appreciation of composers' and performers' pleasing choices. Through the prism of evaluation parameters, music, formerly seen as an abstract and mystical kind of art, becomes more approachable and intelligible.

One of the main purposes of music assessment metrics is to facilitate comparisons between different forms of music. Music encompasses a wide range of genres, styles, and traditions from all over the world. Evaluating these varied musical expressions is necessary for a variety of objectives, ranging from determining the best performance in a competition to selecting the appropriate music for a certain setting, such as a film soundtrack or a marketing campaign. Metrics enable us to compare and evaluate the aspects of music, assisting us in determining which genre of music is best suited to certain criteria or preferences. This comparison component of music assessment measures allows us to make informed decisions while also appreciating the great range of musical creation seen across countries and genres. As artists and composers draw inspiration from many traditions and genres, it encourages cross-pollination and creativity within the musical environment.

2. Related Work

Bridging the gap between subjective and objective assessments is difficult. Subjective assessment takes into account the listeners' own tastes and feelings, whereas objective evaluation is based on quantitative and computational analysis. The difficulty is to strike a balance and establish a link between these two approaches. Future study might concentrate on creating ways for combining these two approaches to give a thorough review [6].

It might be difficult to evaluate musical composition using hearing testing. The risks associated with these evaluations include possible discrepancies in the assessment systems. These inconsistencies can appear as a mix-up between inquiries about the artistic merits of a musical composition and inquiries about the authorship of the composition, whether it was written by a human or an AI [7].

Both professional and novice musicians have successfully integrated the models, the authors have created into their musical endeavours. The web interface implemented on folkrnn.org offers a more user-friendly approach compared to directly executing computer code. Nonetheless, as they continue to gain insights into users' preferences and needs, our ongoing objective is to enhance the overall usability of our system. The researchers observe clear indications of users engaging in an interactive quest for creative inspiration through the iterative cycle of generating and adjusting parameters, exemplifying the system's capacity to stimulate creative exploration [8]. It is unclear if machines can develop the requisite behavioral goals on their own to support artistic performances.

However, when it comes to autonomy in the context of musical creation, a distinct line might be drawn. After being properly educated with expert curated material, modern AI music generators exhibit the potential to create musical compositions on their own without needing constant input from programmers [9]. Identifying the musical preferences of the users is the only requirement for these systems Scale consistency, tone distribution, consecutive pitch repetition, rhythm frequency proficiency, rhythm variability, and other important metrics are used to assess pitch and rhythm features. A variety of musicological objective assessment metrics are incorporated into the modern approach to metric design, providing a way to evaluate and contrast the output produced by musical generation models [10]. These measures have undergone empirical experimentation to validate them, and the findings show reproducibility.

3. Dataset

The dataset was for a type of classical piano music that was influenced by Bach, Mozart, Beethoven, Schubert, Dvorak, Faure, Haydn, Ravel, Brahms, and Cavini. The decision to use a classical piano dataset in the search of sound research has a considerable impact on the results and the level of quality of the conclusions. We selected to use this classical music dataset for a number of compelling reasons -

Depth and completeness of the data, the dataset includes several MIDI (Musical Instrument Digital Interface) files for each of the 10 great classical composers and 330 MIDI files in total. These labels include details regarding the timing of each note, the instrument that plays it, and the note's place within the metrical structure of the musical composition. These annotations are produced using a technique called dynamic temporal warping, which aligns musical notation with audio.

This ensures that the model will not underfit due to the lack of data. Primary reason to use MIDI sequencing files over audio was Compactness and Interpretability. MIDI is an industry-standard format which supports various digital audio workstations and using a MIDI data set would ensure that the dynamics, harmonic and rhythmic information is extractable more conveniently than using audio data. MusicNet, a site that provides a comprehensive diverse library of classical music, is made possible by the cooperation of numerous sources, including the Isabella Stewart Gardner Museum, the European Archive, and Musopen. It was created with assistance from the CIFAR initiative "Learning in Machines and Brains" and the Washington Research Foundation Fund for Innovation in Data-Intensive Discovery.

4. Data Preprocessing

In the proposed work, Visualization is performed on the notes in MIDI samples using the music21 library as part of our data processing methodology. To find anomalies, notes' distribution were plotted, computed their average frequency, and built a count dictionary to identify unique notes. To ensure a streamlined dataset, a list is created for unusual notes (occurring less than 100 times) and removed them. This thorough preparation improves the quality of the dataset for later research phases and lays the groundwork for in-depth analysis.

5. Proposed Work

The proposed work is focused to provide the metrics which helps to identify the AI generated Music and the Humans. Also, discussed the major lacking points of the AI generated music in the metrics.

We have chosen to use Long Short Term Memory (LSTM) models rather than Recurrent Neural Networks (RNNs) in our approach to music production. This decision was made since it was determined that LSTM models were more effective for the task. LSTM models perform better in these areas than RNNs, which have difficulty acquiring long-range relationships and keeping contextual information inside musical sequences [16].

Understanding that music frequently incorporates complex patterns and structures that extend across time led to the choice of using LSTM models in music creation. RNNs frequently experience the vanishing gradient problem, which makes it difficult for them to accurately detect these long-term relationships. LSTMs are more proficient at retaining the contextual details and dependencies inherent in musical compositions because they have memory cells that enable them to remember and use information from earlier time steps.

To improve the quality and consistency of the created music by utilising LSTM models with MIDI files, guaranteeing that it catches the delicate interaction of musical parts across lengthy periods. The advantage of making use of LSTM for the generation of music is listed below -

1) LSTM excels at retaining complex relationships and functional harmony in musical compositions, resulting in more relevant and harmonically rich music generation.

2) LSTM effectively addresses the vanishing gradient problem, which is a limitation of standard RNNs, ensuring better learning of long-term dependencies in music sequences.

3) LSTM networks offer superior parallelization capabilities during training, significantly speeding up the training process. This allows for the exploration of larger and more complex models, leading to improved generative capabilities and the production of high-quality music.

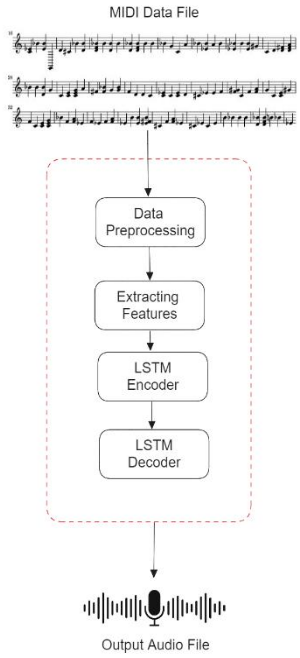

Fig. 1. Low Level Architecture of predicting the notes from the MIDI File using LSTM

The process to train the model on the MIDI file to generate the music using LSTM is provided in the Fig 1. The data processing is performed on the MIDI file like normalization and other processing task to prepare the file. Then the features are extracted. After that, the features are provided to the LSTM Encoder and Decoder. Then the final Output Audio File is generated.

In the implementation of our research work, the essential library in Python played a central role, providing a comprehensive set of functions for efficient audio analysis and processing. Because of its adaptability, feature extraction was done smoothly and fine-tuned parameters could be adjusted. Although we also utilized librosa for extra analysis, the library's efficiency made it the first pick. The wide range of capability offered by Essentia was essential in guaranteeing precise and efficient outcomes for our analysis of audio data.

5.1. Pitch Contour

First, The features like pitch onset, melodic contour and note duration can be measured through a pitch contour analysis. This is essential because a human made melody would contain several subdivisions with some ebb and flow in the melody to invoke the desired emotion. This is to check if all the notes of the melody generated are of the same length thereby making it robotic without any emotional intent. Pitch onset detection can be done with pitch contour analysis. This refers to the point at which a musical note or speech sound begins. It is usually measured in terms of the starting frequency.

Melodic Contour can be extracted with Pitch Contour analysis to determine the shape or pattern formed by the pitch changes in a melody. It can be ascending, descending, cascading, undulating or have various other shapes, which contribute significantly to the emotional intent of the music.

Note Duration indicates how long a particular note or sound is held. This parameter can also impact the emotional expression in the music. Long sustained notes may be played to have more intervallic expression compared to staccato ones. Note Duration is also essential in determining the structure of a song(verse, chorus, etudes etc).

5.2. Amplitude Envelope

This is to see if dynamically all the notes have the same value, this is a key feature that separates human made music from an AI. The shift in the velocity for intended notes is done to put emphasis on that particular interval to invoke a desired emotion. If the true peak of all the transients are the same that means there is no dynamicity in the ai generated melody.

Human made music often exhibits a wide dynamic range. This means that there are variations in volume throughout the composition, which can enhance the emotional impact and create a more natural and engaging listening experience. AI generated music should be able to mimic this dynamic range. If all notes have the same consistent amplitude throughout the composition, it may sound monotonous and lack the expressiveness and emotional intent in human made music.

Having a consistent peak amplitude for all transients in AI generated music could result in a lack of articulation and realism. Human-made music often features nuanced variations in transient amplitudes to simulate the natural sound of instruments.

An audio engineer evaluating AI-generated music will pay close attention to the amplitude envelope to determine if it exhibits the desired dynamics, emotional expressiveness and articulations. Emulating this can help bridge the gap between AI-generated music and music created by composers.

5.3. RMS Energy

The evaluation of the AI-generated tunes obtained from training on a classical piano dataset relies heavily on RMS Energy analysis. This study is important for evaluating the composition's energy changes over time and its melodic development. Likewise, it is crucial in determining the structural components of the resulting work, such as etudes.

In the field of music creation, audio engineers find the examination of RMS energy to be especially useful. Engineers can successfully maintain a pleasant and well-balanced mix by closely examining the RMS levels of the AI-generated piano melody in relation to other musical instruments or compositional elements. This analysis gives them the ability to change the mix as needed, ensuring that no one instrument or section dominates the others and maintaining the desired musical aesthetics.

Etudes are compositions in the classical music genre that are heavily focused on exercises and technical obstacles. Finding particular sections of the AI-generated tune that can present performers with technical difficulties can be done by looking at the RMS energy. With the use of this important information, engineers may work with composers to improve key passages, ensuring that they not only fit the musical concept but are also technically feasible, improving the composition's overall quality and performance.

5.4. Zero Crossing Rate

This gives us a rough indicator of noise and pitch variations. More the zero crossing rate higher the pitch and vice versa.

In Pitch Variation, ZCR can provide a rough indicator of pitch variation within an audio signal. In general, audio signals with a higher pitch tend to have a higher ZCR because the waveform crosses the zero point more frequently due to the rapid oscillations associated with high-pitched sounds. The basic tool for assessing the pitch characteristics of audio signals, although more sophisticated pitch analysis methods are often employed for precise pitch determination.

5.5. Chromogram

The classical piece should fill out the whole spectral frequency ranging from the lowest octave C0 to the highest one C6. A chromagram will aid in the visual representation of the audio generated thereby showing if the piece sounds complete or a subset of the whole piece.

For classical music or any genre that requires a broad range of frequencies, a chromagram can help verify if the composition adequately covers the entire frequency spectrum, from the lowest octave[C0] to the highest one[C6].This ensures that no crucial musical elements are missing, especially in classical music where various instruments cover a wide range.

For Completeness Assessment, a chromagram can serve as a visual representation of the tonal content of a piece of music. If some pitch classes are consistently absent or underrepresented, it may indicate that the composition is not filling out the entire spectrum as expected.

In classical music, different instruments and sections of the orchestra play specific roles, each contributing to the overall sonic spectrum. An analysis of the chromagram can reveal which instruments or sections are contributing to which pitch classes. An audio engineer can compare the chromagram of AI-generated music to that of reference pieces composed by humans. This comparison can reveal differences in tonal distribution, harmonic complexity, and completeness, helping to fine-tune the AI-generated music for better alignment with human-created works.

6. Result and Discussion

In order to determine whether AI-generated music has the dynamics and quality of a human-generated melody, we will attempt to identify the AI-generated music based on the metrics. Also, to determine how the generated melody might be improved upon and identified, we have compared it to a human-made melody using graphs and plots, utilizing the metrics included in the proposed approach.

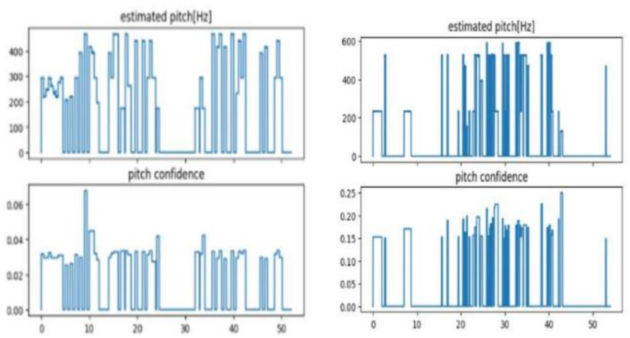

Fig. 2. Pitch Contour Analysis on the AI generated music (Left) and the Human generated music(Right)

Fig. 2, shows a thorough examination of the AI-generated music file's pitch contours, whereas the second shows a comparable analysis of a piece of music written by a human. Notably, when contrasting these two approaches, a crucial discrepancy becomes apparent.

The pitch contour in the AI-generated music file shows some visual unpredictability in the flow of the pitch. This contrast draws attention to the distinctiveness of AI-generated compositions, where the melodic structure may differ from those created by humans. The accuracy of the melody contour, as seen in Figure 1, highlights the AI's ability to discover unusual musical pathways.

The human-composed music file's pitch contour analysis demonstrates that the melodic transitions have a better level of consistency and coherence. The human-composed audio file's closeness of contours shows a planned and organized change from one part of the composition to another. This feature emphasizes how human-composed music has a unique structural character and how transitions are frequently influenced by creative purpose and conventional music practices.

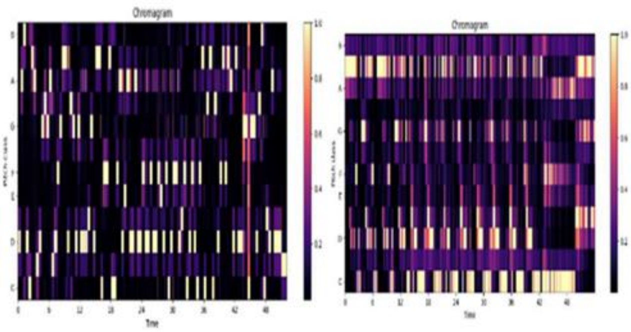

Fig. 3. Chromagram results on the AI generated music (Left) and the Human generated music (Right)

Figure 3 shows a chromagram of an AI-generated music file, whereas Figure 4 shows a chromagram of a piece of music composed by a human composer. When contrasting these two representations, a noteworthy and distinctive feature becomes apparent.

The broad covering of the frequency spectrum is a prominent aspect of the chromagram in the case of the humancomposed music file, as shown in figure 4. Throughout the whole file, this coverage remains constant. This feature emphasizes how human composers are capable of writing music that makes use of a wide variety of frequencies, adding to a full and varied aural experience.

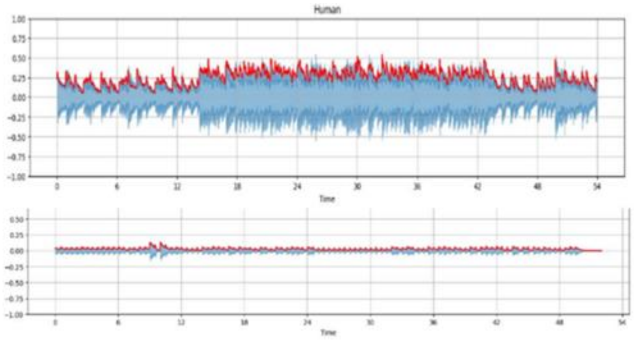

Fig. 4. Amplitude Envelope result on the AI generated music (Bottom) and the Human generated music (Top)

The human-composed music file's potential to demonstrate a greater dynamic range and amplitude variability is clearly demonstrated by the amplitude envelope analysis. In Fig. 4., the AI-generated audio file, on the other hand, displays a striking consistency in amplitude, indicating the need for a more thorough training method that takes into account a wider range of characteristics.

The diverse velocity patterns that may express emotional intent and preserve contextual relevance across multiple musical styles and genres must be included in order to give the AI-generated compositions more expressiveness and emotional resonance. The apparent absence of melodic movement and structural coherence in the AI-generated track highlights its existing limits and highlights the need for more enhancement and diversity in the training approach and dataset.

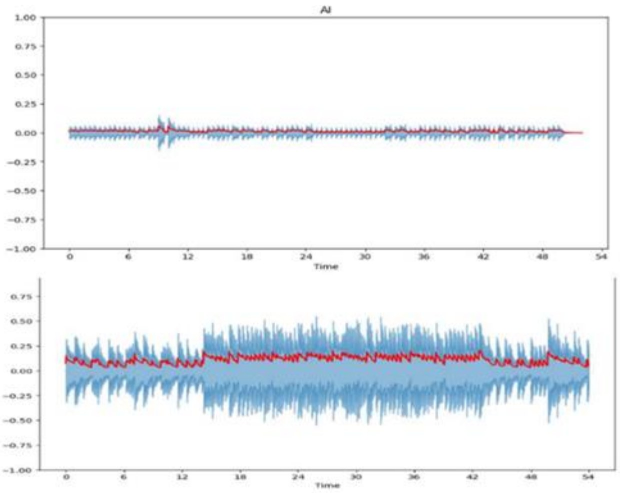

Fig. 5. RMS energy result on the AI generated music (Top) and Human generated music (Bottom)

In Fig. 5, The density in the rms energy for the human composed music is varied and noticeable while there isn't any much difference in the energy levels of an ai generated music piece. This helps us evaluate the lack of excitement generated by the AI. Comparing RMS energy can provide insights into the dynamic range of the audio. A larger difference in RMS energy between the quietest and loudest parts of an audio file indicates a wider dynamic range.

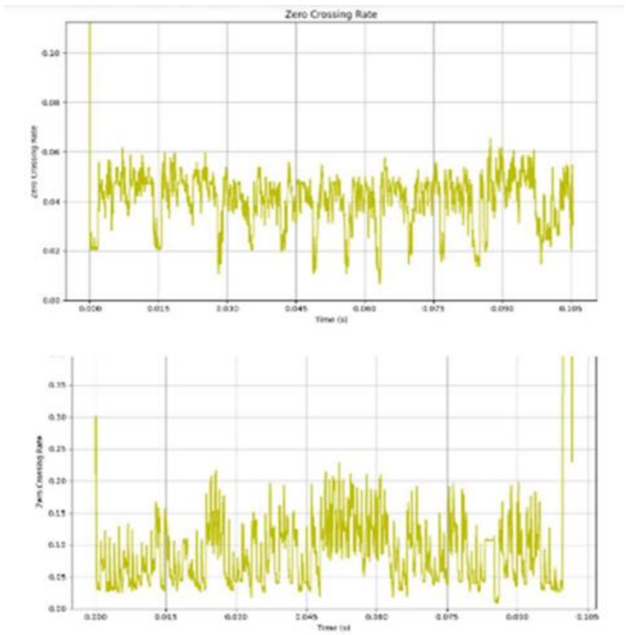

Fig. 6. Zero Crossing Rate results on the AI generated music (Top) and Human generated music (Bottom)

In Fig. 6, we can observe values of ZCR in the human composed music are wider in range compared to the ai generated piece indicating more complexity and variation. Human composed music has more discernible motifs and recurring patterns making the structure more memorable while ai generated music lacks it. Zero crossing rate serves as a basic descriptor in music analysis, offering insights into aspects of timbre, transients, and potential genre characteristics.

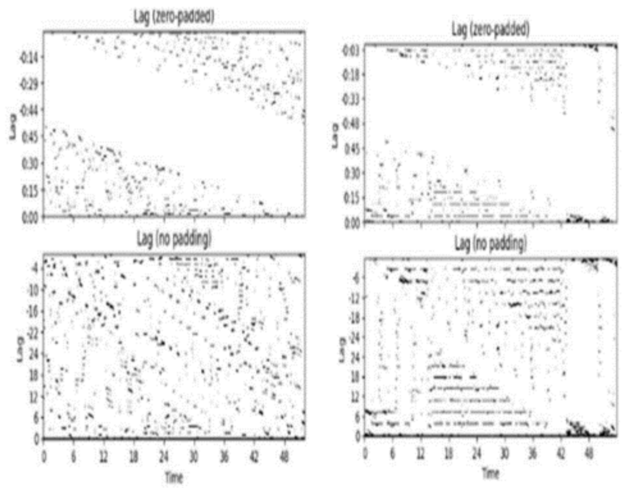

Fig. 7. Temporal Frequency results on the AI generated music (Left) and the Human generated music (Right)

In Fig.7, The temporal frequency is a measure of how quickly events occur in time, and in the context of music, it corresponds to the perceived beat or tempo. Higher points on the y-axis indicate higher temporal frequencies or faster beats, while lower points represent slower beats. In complex music pieces, you may observe multiple peaks or lines at various temporal frequencies in the tempogram. These multiple peaks can indicate the presence of multiple rhythmic patterns or tempo changes within the music.

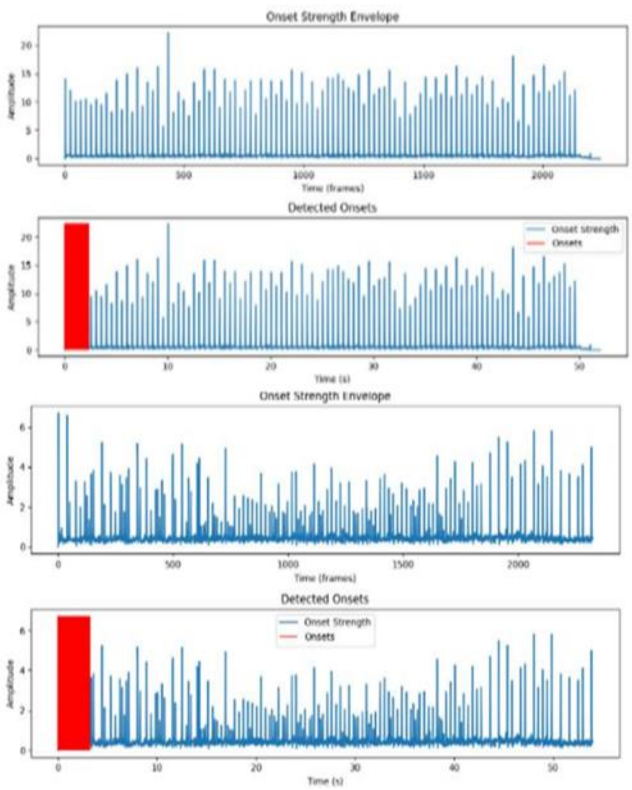

Fig. 8. Recurrence matrix results on the AI generated music (Top) and the Human generated music (Bottom)

In Fig. 8., lag_pad, denoting the lag representation with padding, incorporates zero-padding to maintain uniform dimensions across all elements, while lag_nopad, representing the lag without padding, may exhibit shorter dimensions. These lag representations serve a valuable purpose in discerning temporal connections within musical events, facilitating the analysis of patterns' repetition intervals. Notably, this allows us to distinguish the dispersion of patterns in AI-generated music, in contrast to the recurring motifs observed in compositions by human musicians.

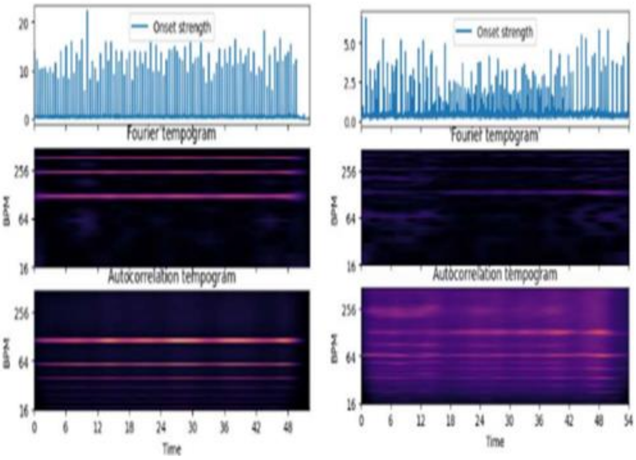

Fig. 9. Onset Detection/ True peak results on the AI generated music (Left) and the Human generated music (R)

In order to identify the temporal occurrences that signify major musical events or note attacks, onset detection techniques are essential. In comparison to AI-generated compositions, human-composed music has a unique advantage due to the frequency of onsets and the subdivisions of these onsets. In Fig.9., the temporal intervals between subsequent onsets are where this difference in rhythmic intensity is most noticeable. Human-created music tends to have a more complex and dynamic rhythmic structure, whereas AI-generated music tends to have a subdued rhythmic profile.

7. Conclusion

This research has shed light on the fascinating area of AI generated music and has undertaken a critical and thorough examination of its limitations. Through a comprehensive metrics evaluation, we have discerned several key insights. We used features like chromagram, Amplitude Envelope, Zero Crossing Rate and RootMeanSquare Energy to determine the dynamicity, completeness and unpredictability of AI generated audio file and how those aspects can be more elevated and made nuanced. The evaluation metrics above form a better basis for judgement while comparing the competence and richness of an AI generated music sample compared to the traditional evaluation metrics which involves a lot of mathematics and less graphical representation to make the required deductions. The idea that "the little imperfections are where the creativity lies" in music refers to the belief that it's often the subtle nuances, quirks, and deviations from perfection that make a piece of music uniquely expressive and captivating.

However, it is important to acknowledge the strides that AI has made in democratising music production, offering new avenues for collaboration between humans and machines, and serving as a source of inspiration for human artists. The integration of AI tools in the creative process is an exciting and evolving area that holds great potential.

8. Future Scope

The current state of AI-generated music is undeniably promising, showcasing remarkable advancements in machine learning and creativity. However, it still falls short in replicating several crucial facets of professional musicianship. A noticeable limitation lies in the lack of semantic context over longer musical durations. AI-generated compositions often lack the repetition of motifs or licks that serve as signature elements in making a musical piece truly memorable. What sets human music apart is its dynamic and unquantized nature, which imparts a sense of realism and liveliness. Human musicians make specific intervallic choices to evoke emotions, creating a profound connection between music and the listener. This subtlety of emotion-driven music composition remains a complex challenge for AI systems to master fully.

Author contributions

Hrishikesh Yadav: Technical documentation, Experimentation and Writing-Reviewing. **Prerak Joshi:** Experimentation and Implementation of the Approach, Writing. **Jay Oza:** Technical documentation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] M. Hilsdorf, “MusicLM — Has Google Solved AI Music Generation? - Towards Data Science,” Medium, Jun. 11, 2023. https://towardsdatascience.com/musiclm-has-google-solved-ai-musicgeneration-c6859e76bc3c

[2] “MusicGen: Simple and Controllable Music Generation.” https://ai.honu.io/papers/musicgen/

[3] “Open sourcing AudioCraft: Generative AI for audio made simple and available to all,” Meta, Aug. 02, 2023. https://ai.meta.com/blog/audiocraft-musicgen-audiogen-encodecgenerative-ai-audio/ (accessed Nov. 17, 2023).

[4] H. -e. Schaefer, “Music-Evoked Emotions —Current Studies,” Frontiers in Neuroscience, Nov. 24, 2017. https://doi.org/10.3389/fnins.2017.00600

[5] “Why Music Matters: The Cognitive Personalism of Reimer and Elliott.” https://cmed.ku.edu/private/daugherty.htmlXiong, Zeyu, Weitao Wang, Jing Yu, Yue Lin, and Ziyan Wang. "A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music." arXiv preprint arXiv:2308.13736 (2023).

[6] Li-Chia Yang and Alexander Lerch, “On the evaluation of generative models in music”, Neural Computing and Applications 32, 9(2020), 4773–4784.

[7] Oded Ben-Tal, Matthew Tobias Harris, Bob L.T. Sturm, “How Music AI Is Useful: Engagements with Composers, Performers and Audiences”, Leonardo 2021, 54 (5): 510–516. doi: https://doi.org/10.1162/leon_a_01959

[8] Joo-Wha Hong, Katrin Fischer, Yul Ha, Yilei Zeng, “Human, I wrote a song for you: An experiment testing the influence of machines’ attributes on the AI-composed music evaluation”, Computers in Human Behavior, Volume 131, 2022, 107239, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2022.107239.

[9] Xiong, Zeyu, et al. "A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music." arXiv preprint arXiv:2308.13736 (2023).

[10] Hyeshin Chu, Joohee Kim, Seongouk Kim, Hongkyu Lim, Hyunwook Lee, Seungmin Jin, Jongeun Lee, Taehwan Kim, and Sungahn Ko, “An Empirical Study on How People Perceive AI-generated Music”, In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM '22), 2022, Association for Computing Machinery, New York, NY, USA, 304–314. https://doi.org/10.1145/3511808.3557235

[11] D. P. Nicolalde Rodriguez, J. A. Apolinario and L. W. P. Biscainho, "Audio Authenticity: Detecting ENF Discontinuity With High Precision Phase Analysis", In IEEE Transactions on Information Forensics and Security, vol. 5, no. 3, pp. 534-543, Sept. 2010, doi: 10.1109/TIFS.2010.2051270.

[12] Bachu, R. G., et al. "Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal." American Society for Engineering Education (ASEE) zone conference proceedings. American Society for Engineering Education, 2008.

[13] Kauppinen, Ismo, and Kari Roth. "Audio signal extrapolation–theory and applications." Proc. DAFx. 2002.

[14] Luca Angioloni, Tijn Borghuis, Lorenzo Brusci, and Paolo Frasconi, “Conlon: A pseudo-song generator based on a new piano roll, wasserstein autoencoders, and optimal interpolations”, In Proceedings of the 21th International Society for Music Information Retrieval Conference ISMIR MTL2020, 2020, 876--883.

[15] Mohit Dua, Rohit Yadav, Divya Mamgai, Sonali Brodiya, “An Improved RNN-LSTM based Novel Approach for Sheet Music Generation”, Procedia Computer Science, Volume 171, 2020, Pages 465-474, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.04.049.

[16] E. Dervakos, G. Filandrianos and G. Stamou, "Heuristics for Evaluation of AI Generated Music," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp.9164-9171, doi: 10.1109/ICPR48806.2021.9413310.

[17] Bogdanov D, Wack N, Gómez E, Sankalp G, Herrera P, Mayor O, Roma G, Salamon J, Zapata J, Serra X, “Essentia: an audio analysis library for music information retrieval”, In: Britto A, Gouyon F, Dixon S, editors. 14th Conference of the International Society for Music Information Retrieval (ISMIR); 2013 Nov 4-8; Curitiba, Brazil. [place unknown]: ISMIR; 2013. p. 493-8.

[18] M. F. McKinney and J. Breebaart, “Features for audio and music classification.,” in ISMIR, 2003, vol. 3, pp. 151–158.

[19] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, “The million song dataset,” in ISMIR 2011: Proceedings of the 12th International Society for Music Information Retrieval Conference, October 24-28, 2011, Miami, Florida. University of Miami, 2011, pp. 591–596.

[20] McFee, Brian, et al. "librosa: Audio and music signal analysis in python."