DOI: 10.2478/seeur-2024-0075

COMPARATIVE SYMBOLIC ANALYSIS OF THE ETHNO-FUSION GENRE: INSIGHTS AND PERSPECTIVES

Valton Kamberaj, PhD (c)
Faculty of Contemporary Sciences and Technologies,
South East European University,
Tetovo, North Macedonia
vk30443@seeu.edu.mk

Arbana Kadriu, Full Professor,
Faculty of Contemporary Sciences and Technologies,
South East European University,
Tetovo, North Macedonia
a.kadriu@seeu.edu.mk

Nuhi Besimi, Assistant Professor,
Faculty of Contemporary Sciences and Technologies,
South East European University,
Tetovo, North Macedonia
n.besimi@seeu.edu.mk

ABSTRACT

This study explores the integration of music and technology, illustrating their potential to collaboratively push the boundaries of musical exploration. Despite traditionally being viewed as unrelated, the combination of these two fields can significantly contribute to the progress of musical development. This study uses advanced computational methods to build a dataset filled with symbolic musical sequences that belong to a specific genre. This dataset is shown to be highly accurate and provides a detailed analysis of frequencies when examined closely, highlighting its quality and depth. We subject our dataset to comparative analysis with the renowned MAESTRO dataset, employing chromagrams to examine audio signals, rhythms, chords, solos, and note patterns in MIDI format through a variety of methods. This comparison underscores the superior quality of our sequences relative to those in the MAESTRO dataset, emphasizing the meticulousness of our sequence creation process. Moreover, we conduct internal evaluations of our dataset using both three-dimensional and two-dimensional approaches to melody representation, confirming its viability for future scholarly work. This effort seeks to enhance the music field by integrating computer science insights and methodologies, expanding the scope for future music technology research. It highlights the collaborative potential between musical creativity and technological advances in ongoing studies.

Keywords: MIDI, Dataset, Analysis, Frequencies, Music, Melodic.

INTRODUCTION

Computer science, as one of the most important fields today, has many other connecting and cooperating parts with which you can help create products that are very important for humanity. A very important connection is also with the field of music, which plays a very big role in our lives. Considering many points where music is expressed through sounds at different frequencies and combinations, can give a very important signal to the human ear (Martineau 2008; Silla et al. 2017).

Musicians and composers who deal with these frequencies or their combination to create melodic lines that are expressed through notes, create a unique combination referring to a certain genre (Oore et al., 2018).

Considering that music is a field that is generally called art, it is also a very dimensional field, taking into account many points without which it is not able to function and be called music (Peterson et al., 2017). The dimensions without which a certain melodic line cannot be expressed and cannot have a basic meaning are sound, timbre, and rhythm (Anand et al., 2020). To give a basic meaning to a line you must have a solo, chords, and beat. These three sub-dimensions are essential, they give the basic meaning, which can then be expressed through other dimensions; the more dimensions are included in that combination of frequencies, the more accurate and understandable it becomes by calling it complete (Hawthorne et al. 2021; Chen et al. 2020).

Various techniques are used to achieve the accuracy of these dimensions, especially these basic ones, without which it is impossible to properly express the meaning of that melody. The accuracy of the note and the timbre of the frequency is the basic part, without which one cannot continue further by weaving it with precision into a unit of time, which is called rhythm (Gatti et al. 2017; Briot 2021; Silla et al. 2017).

The creation of accurate melodic lines is the basis for continuing with experiments and the use of different algorithms for the purpose of generation or composition or other analyses that are intertwined in these two fields.

In this paper, we present an analysis aimed at understanding the importance of creating accurate notes within a unit of time and comparing them with other melodic creations by different composers, well executed by different instrumentalists. The topic is introduced with the construction of an initial corpus with melodic lines that correspond to a unique genre. The analysis is based on MIDI, which according to different authors is described as the most efficient format for algorithms to work with, producing accurate and faster results. It is essential to note that when working with MIDI, we are dealing with vector or symbolic data that consists of only numbers with a specific frequency, which are readily inserted into computer programs. Therefore, we are no longer working with "music" but with data.

RELATED WORK

Creating sequential symbolic notes is essential for accurate analysis and testing of audio data. The authors (Gómez et al., 2015) extended their work on constructing guitar sequences in the Flamenco genre. Their tests focused on modeling the frequencies of vocal contour and pitch, yielding results of average accuracy. However, they encountered noticeable challenges due to noise in the sequences. Another analysis based on musical dimensions was presented by Kentaro Shibataa et al. (2021). In their work, the system focused primarily on two dimensions: multi-pitch detection and rhythm quantization. The error rate in the processing and results is estimated to be 7.1%. The system effectively maintains the regular order of note frequencies and accurately places them within the rhythm. The dataset created in MIDI is highly important for various analysis purposes, with symbolic sequences being the most utilized and combined. This facilitates easier analysis, testing, and deriving results from different computer and mathematical models (Qiuqiang Kong et al., 2022; Kong et al., 2021; Nakamura et al., 2017). Considering various analyses from numerous works and authors, the use of different datasets is consistently based on the analysis of sequences constructed with MIDI. MIDI is one of the primary tools that generates the most accurate frequencies compared to other instruments (DuBreuil, 2020; Sarmento et al., 2023; Qiu et al., 2023; Colton, 2021). Other analyses with piano datasets have been done, which, taking into account the chromagrams and constructions of pitch and note detection, have obtained good results considering that the sequential construction of notes from other datasets is correct (Shih-Lun Wu et al. 2022; Xianchao Wu et al. 2020; Sreetama Mukherjee et al. 2022). The analysis results with other sequential construction instruments were not achieved in other works (Sarmento et al., 2023; I. Barbancho et al., 2012; Yu-Hua Chen et al., 2020), highlighting the difficulty in detecting sequential grades without errors. Instruments like the guitar are used much less frequently because their analysis results are not as desirable as those generated with piano-roll (Chris Donahue et al., 2019).

DATA COLLECTION

The data we work with are MIDI files that make up our dataset and were created with the help of music professionals. Analyses and comparisons are conducted using one of the largest datasets of piano-roll MIDI sequences in mixed classical music, known as MAESTRO. This dataset comprises approximately 200 hours of melodic lines (Curtis Hawthorne et al., 2019). The sequences in our dataset were manually created using programs such as FL and Logic, where the main goal was the accuracy of the frequency construction with this method. We present all the results of the comparisons between our dataset and MAESTRO in the "experimental results" section.

In this study, we will delve deeper into the symbolic or MIDI form of corpus construction, focusing on data built for a specific genre with a unique combination of frequencies. This symbolic structure was chosen for the melodic line building because it is the best, most accurate, and most understandable format for construction. The symbolic sequences can demonstrate the actual precision of the note in that each "staccato" - a division of the frequency that expresses tonality and has a distinct meaning. It has been decided to construct the first corpus in this genre that contains melodic lines as a feature lacking in other weaved genres in MIDI format. Three fundamental dimensions are used in our corpus: solo, chords, and rhythm. These dimensions are sufficient for both music professionals and listeners to understand the meaning of a melodic line. We have performed one thousand melodic lines in our genre under the guidance of expert musicians, all of which have been expressed with remarkable accuracy. Every note in the melody has been the subject of additional observations to produce the melody with the highest accuracy possible in terms of identifying the notes in time and rhythm, without any noise expressed as frequency, because everything was done directly and symbolically from the start.

EXPERIMENTAL RESULTS

During the experiments, we see the results obtained through different analyses, determining the importance of sequences, their placement, and timing in the most accurate way, and then comparing the results with the large dataset of MAESTRO. The experiment starts with the chromagram, where both our dataset in the form of an audio melodic line, and another one from MAESTRO's were used as inputs. The other part includes the analysis of chords and beats for our genre in two melodic lines and one of MAESTRO's. The analysis was carried out between two lines in our dataset and then also with that of MAESTRO. The rest of the analysis is done with a note-histogram for all the lines in our and MAESTRO's dataset. Finally, the study’s findings were presented which show the clear difference in accuracy from our dataset.

Chromagram Audio Analysis

The use of audio as input for analysis is done using the chromagram, which gives signals as output of the use of frequencies per unit of time and showing the maintenance of a frequency during a unit of time (N. Rosmawarni T. T et al. 2023; N. Rosmawarni T. T et al. 2023). In this chapter, we will present the experiments done and expressed graphically. Chromagram is the main part of testing and comparing the main dimensions for audio where the main purpose is the comparison between our dataset and Maestro expressed in frequencies.

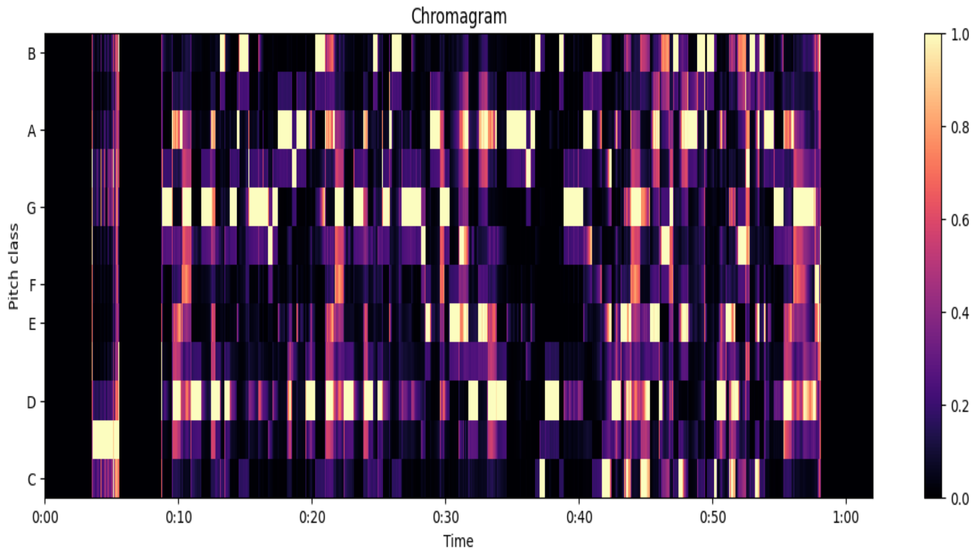


Figure 1. Chromagram analysis-audio signal in ethno-fusion genre.

When we use a technique called the Gaussian method for our first analysis, we focus on something known as a chromagram. This is shown in the first figure. In this chromagram, we can see the musical notes and how they change over time for a piece of music, which in this case is a melody named 'Hymn', which is typical for our music genre.

In Figure 1, the left side represents the main notes that are used for this melodic line. Below it, we notice the time and the beat, which are divided into equal parts to express the speed of the beat of this line, which is a minute in length. On the right is the length of holding a note, noting the point 0.0 which in this case the color "black" shows that that note in that time relation has not been played, and then going up to the point 1.0, to show the length of that note in that time dependence that is kept the most.

When analyzing audio signals, it is important to know that no graph can perfectly represent every sound from the music of any genre. This is because background noise, which we capture alongside the music, might show up in our charts as if it was actual musical notes. In the upper part, even though with an audio signal, we can see a pretty accurate sequence of notes and their length concerning the beat, which is a comparative value in the case when we analyze a melodic line from the MAESTRO dataset.

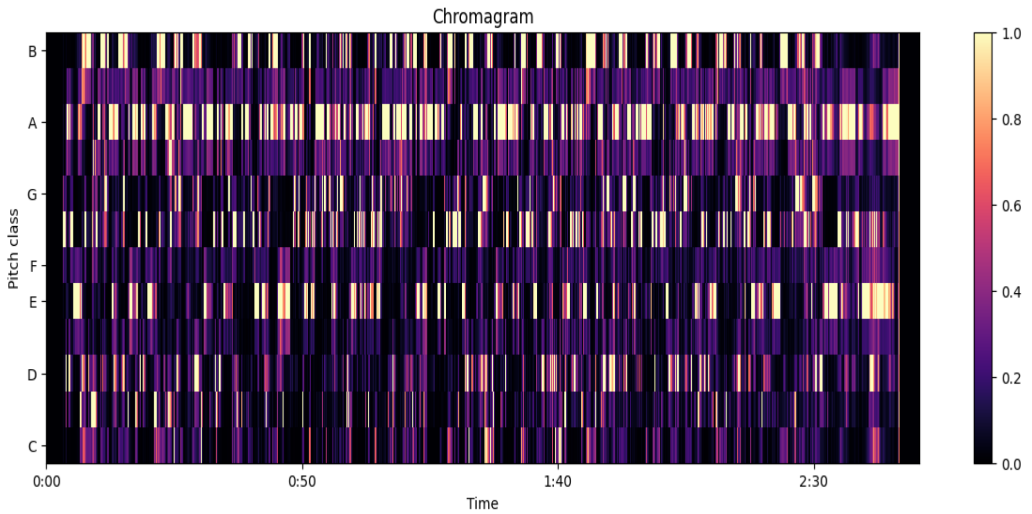


Figure 2. Chromagram analysis-audio signal from MAESTRO dataset.

In our analysis, we use a chromagram to study how loud each note is during a specific time or beat. This helps us understand which notes are played together and how they relate to each other in terms of frequency—whether they are close to or far from one another. We are looking at the music 'Canon in D' by Pachelbel, from the MAESTRO dataset, which is played in the D key .

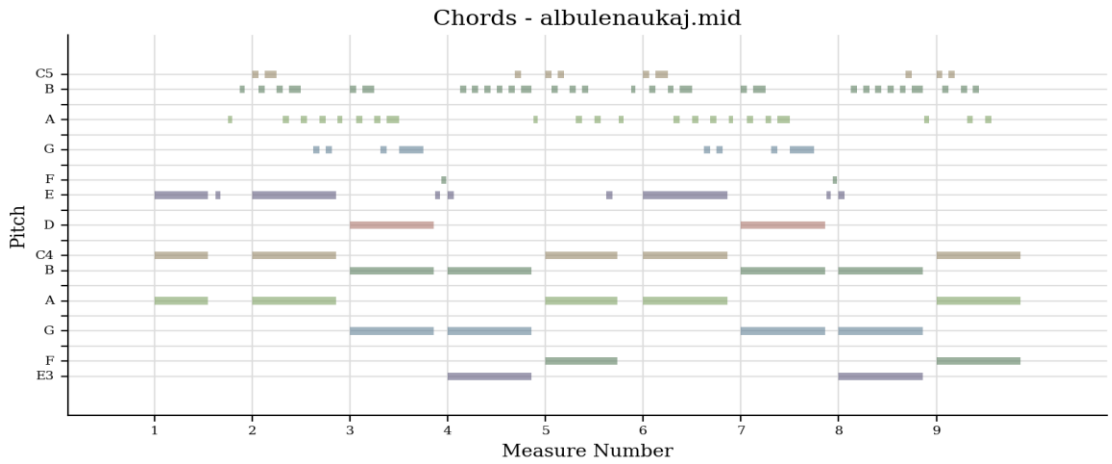
Based on the frequencies of the analysis shown in Figure 2, we can see a sequence of main notes within those frequencies. Given that our genre includes more oriental elements and employs connecting additions that sometimes differ from other music genres, we will account for these variations in our analysis. Due to this, the “legato” and “Arabic scale” parts are missing, with which the use and extension of the notes become inseparable, they are presented as the length of a note which changes gradually to connect two different frequencies. The figure illustrating this scale shows that the note D# is rarely used. However, if a melodic line in the note D existed within our genre, D# would be significantly more prominent, with its intensity values on the right side being at least half or higher.

The division of notes made in the classical genre or MAESTRO's melodic lines is mostly done with 'staccato' and therefore from the figure we see more separated sequences of frequencies represented in units of time, leading to the design and expression of these notes close to the MIDI format. In contrast, it is much more difficult to resemble the separate notes of our genre using the MIDI format.

On the other hand, we have observed that many notes have been recorded as noise in the audio signal, which has resulted in great inaccuracies in the beat. We have also noticed that notes which require intensity have been expressed as black color, this is something that is not observed in our genre even though it is in an audio signal.

Solo, Chords, and Beats Analysis

The figures below illustrate the solo, chords, and rhythm as the three fundamental dimensions in each melodic line. By examining the sequence and timing of the soloist sequences, we can accurately divide the measure into correct parts based on the symbolic frequencies provided by the two datasets. The primary reason for analyzing these dimensions is their representation in the most accurate note frequency, MIDI, which is the core format of our work. From dataset creation to analysis and comparison with other datasets in the same format, MIDI serves as the foundational element.



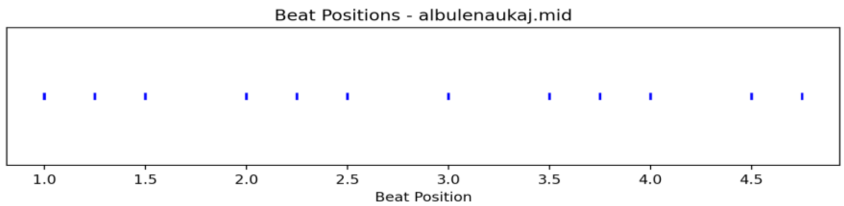


Figure 3. Solo, chords and beats analysis-symbolic sequence in ethno-fusion genre.

In the next analysis, we used the symbolic part, referring to the creation of melodic lines in the most possible accurate way. Creating sounds with musical frequencies to generate a melody by adding it to our dataset is done using the right equipment to generate the three main dimensions that are the basis for giving meaning to that line.

We observe that the sounds are from "albulenaukaj," an artist in our genre. Our team generated melodic lines for this artist, and then recreated them from scratch in symbolic form or MIDI. According to the literature review, using piano sounds yields the most accurate generation and results. In the figure above, we see the three main dimensions which are separate or solo frequencies, the part of the chords that are formed at least by combining three notes, and the tact or the part of the tempo to see the great accuracy of the placement of the notes. Our analysis using MIDI signals reveals the specific positions and notes touched in this unique genre. Each sequence is precisely placed within the given tempo. The chord sections (long sequences) are maintained with high accuracy, matching the tempo exactly. The divisions are clearly expressed below with numbers: 1, 2, 3, etc.

The melodic line played in A minor, which is also the main tone in our genre with a tempo speed of 130 beats per minute and in the "Arabic" rhythm, has used the note B the most, which can be expressed as A#, and that in the analysis of the chromagram we pointed out the importance of the rotation of frequency cases within the genre.

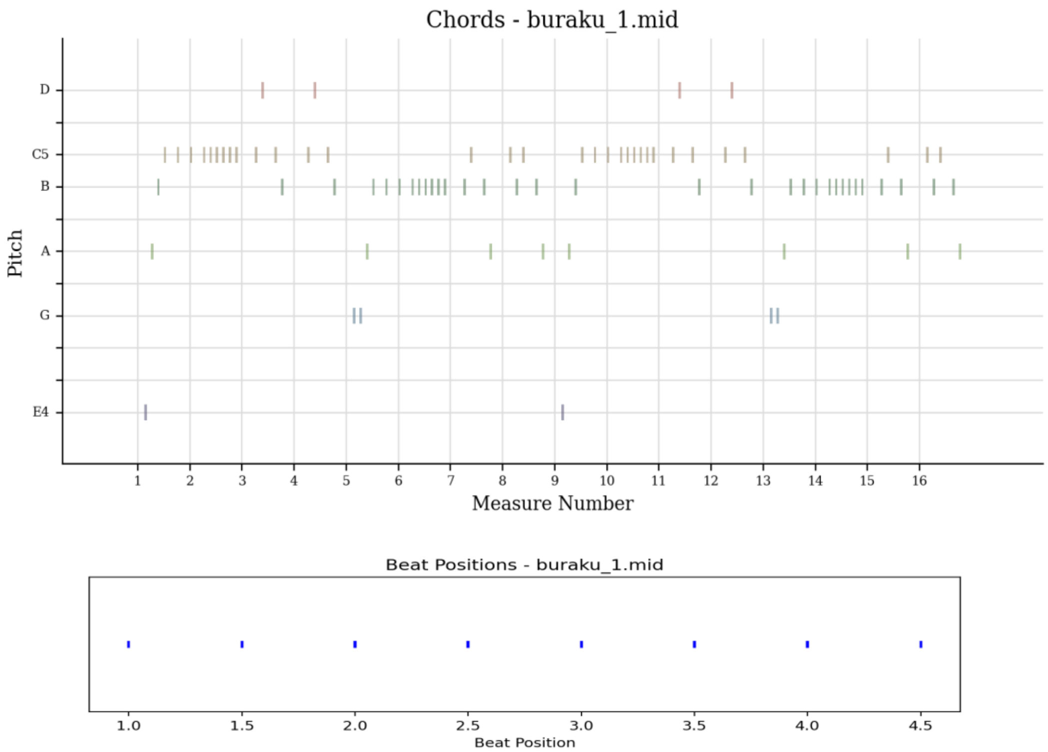
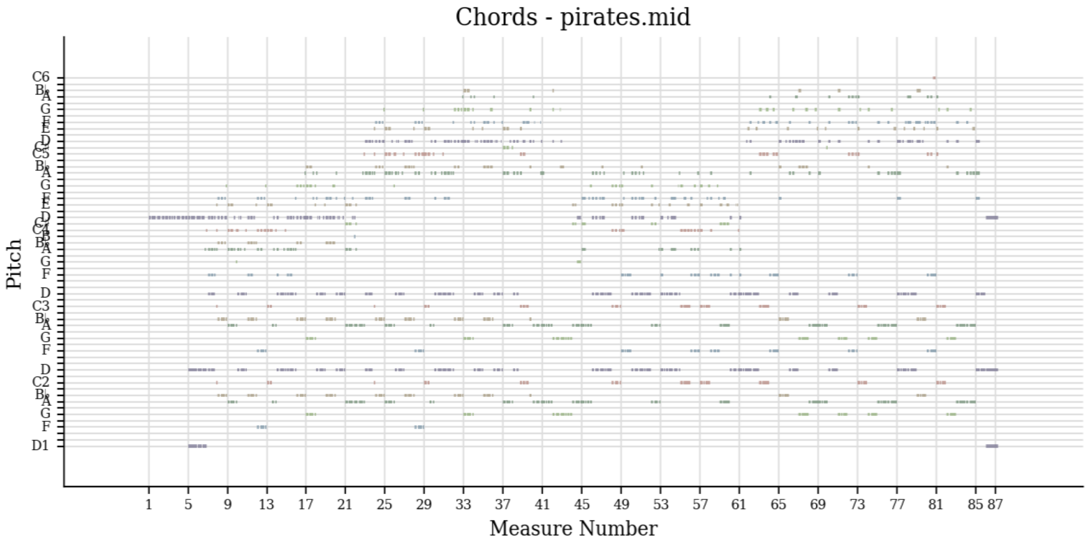


Figure 4. Solo and beats analysis-symbolic (without chords) sequence in the ethno-fusion genre.

From the results of Figure 2, we have an analysis made for some symbolic sequences that we have realized for testing purposes. From the upper part, we notice that initially, the results do not look like the example from our dataset, which was created in a very precise way. The next part of the analysis involves the creation of the melodic line without using the chord part. It is seen how the frequency content is presented only with the solo part (a separate part of the sequences), which is played at the correct time. Still, since the part of the chords is missing, we can say that we have a half-hearted music. It cannot even make sense, resulting in this way that neither for the listener's ear nor for those who analyze the results, this combination of frequencies makes no sense, it is very difficult to see what composers wanted to create. Therefore, this part of the symbolic creation is prohibited in our work. To achieve accurate results, all melodic lines must adhere to the three main dimensions.



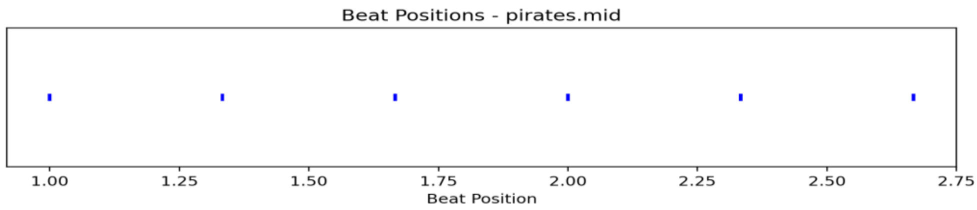


Figure 5. Solo, chords, and beats analysis-symbolic sequence from MAESTRO dataset.

In the analysis of one of MAESTRO's music that we have presented in this paper (we are based on many melodic lines of this dataset), we see a big difference between the frequencies of this genre and our genre. In essence, if the obtained result is analyzed, we will notice an incorrect sequence of notes within the melodic line. This is due to several reasons. Although the three main dimensions—solo, chord, and rhythm—have been considered, it is observed that the rhythm, particularly in the section called "beat position," is disordered. The high notes do not maintain a consistent tempo, resulting in values that change chaotically over different time intervals.

The other part of the analysis reveals that the chords and the solo are not properly synchronized with each other, nor with the beat. According to our analysis, this happened for several reasons. The first is that the transformation in the MIDI sequence is done by taking the external noises in certain places, resulting in an expression in a certain note, for example, if the combination of symbolic frequencies was supposed to be A-C-D, half of them resulted in small sequence such as F# (F with a sharp), which is numbered as a note, confusing the sounds and gradually losing control of the range of frequencies until the end. The second is that the playing of these notes was done by different professionals who have been performing on the piano for a long time, and then the possibility has been very high that in certain places we have small mistakes which, however, are not investigated by the public, it is observed in the community part when we analyze the sequences.

Note Histogram Analysis

In the histogram score analysis part of our genre, we have dealt with all the created melodic lines. The note histogram represents the notes that have been tapped or played on the piano keys, showing how many times note has been played within all those melodic lines (Ali C. Gedik, 2010). The main purpose of this analysis is the correct collection of the used notes expressed in MIDI, which indicate the accuracy of the same notes from which the melodic lines are mostly created in A minor.

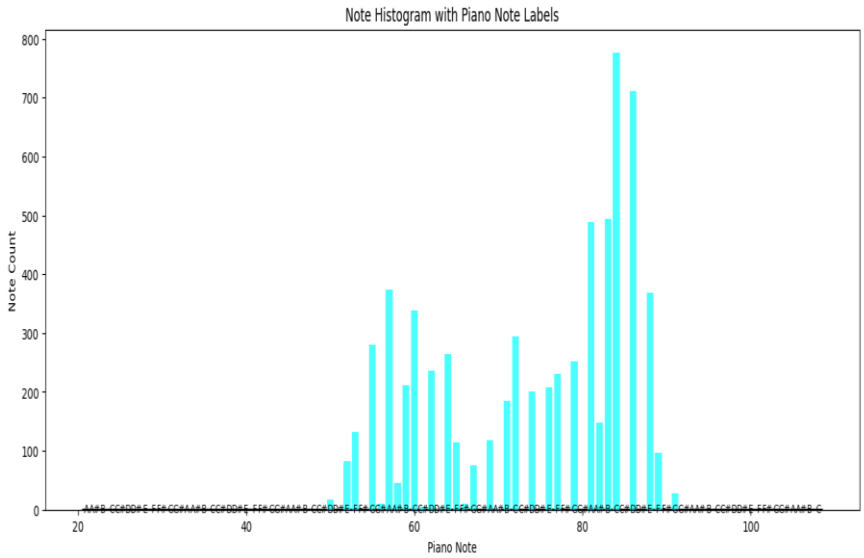


Figure 6. Note histogram in ethno-fusion genre and dataset.

The figure indicates that the most frequently used notes are those leading to A minor, with the note B being particularly prominent. This note is consistently accurate across different octaves.

Another great importance is the collection of submitted grades. In our dataset, 90% of the melodic voices are played in the key of A minor and some even in A major. The tempo used to create the sequences is 120-130 beats per minute. This note histogram shows precisely that the collection of notes from all MIDI files are very close to each other, with this we conclude the great and close accuracy of both the tempo and the melodic lines. This is of great importance because any kind of analysis, generation, or work with this dataset results with greater accuracy and much better results compared to the MAESTRO dataset where we are making the comparisons.

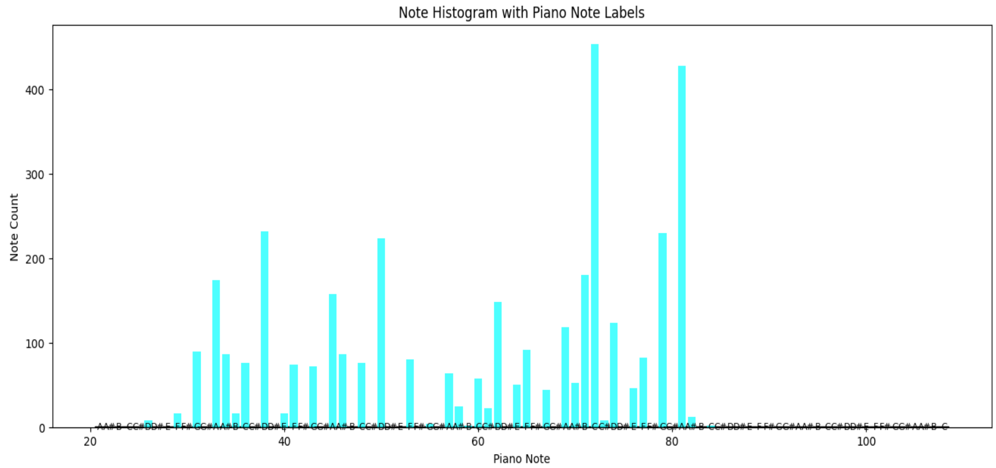


Figure 7. Note the histogram in the MAESTRO dataset.

In the part of the note histogram from MAESTRO's dataset, we have notes that are used in different octaves and tonalities, with which the range of notes is much expanded, compared to our dataset.

This has to do with those frequencies that are classified as noise and it can be seen in the analysis that they are presented in the form of some half notes, but some also in the form of complete sequences. Another reason is that many artists have played numerous octaves and tonalities on the piano, performing works by various composers. Most importantly, the beat structure has continuously changed over time, which is a key factor considered in the analysis above. Well, this part of the beat also comes in here because if a note was supposed to be played 2/4 within a time of 1 second, then if that tempo was not kept and the note exceeded the limit of holding the pressed piano cork, it was read as another note within the next 1 second. The difference is very big in terms of accuracy with our dataset, resulting in future analyses being more accurate in terms of our genre and MAESTRO.

Minor and Major Scale

The other analysis includes an additional dimension: verifying whether the created dataset was built correctly. This analysis supports our claim that a distinguishing feature of our corpus is that while most melodic lines are in the minor scale, the major scale is also prominently represented. This is proven in the sequential accuracy of the notes B and B (minor), which are not combined with D# (sharp), to give the note of the major scale. This is another reason for the accuracy of the dataset creation. Given that many of the world's hit songs are composed in the minor scale, we utilized Key Analysis to verify the accuracy and consistency of this scale in our dataset.

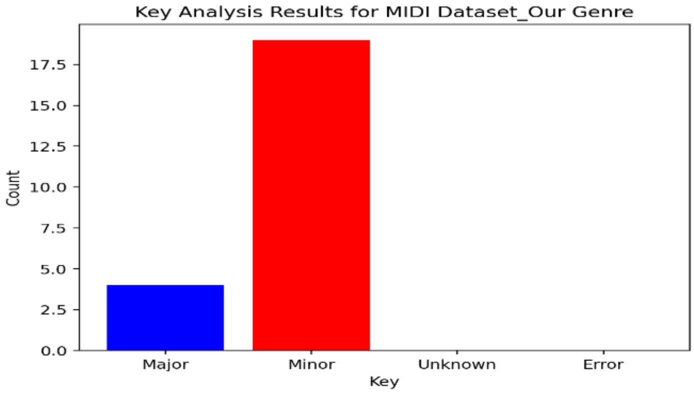


Figure 8. Minor and Major scale.

CONCLUSION

The intersection of computer science and music provides us with numerous opportunities to develop and implement new ideas for enhancing the latter. In this study, we recognize that the workspaces are vast, and making swift improvements or introducing new innovations is challenging due to the complexity and multifaceted nature of noise. Addressing this requires comprehensive efforts and continuous development in the future. Regarding this work, we have brought a group of data in the form of symbolic or MIDI files of a specific genre that is unique, with the aim of sequential analysis and frequencies expressed in MIDI. During the analysis, we noticed different but correct results regarding our melodic lines.

The initial comparative analysis was conducted using a chromagram, comparing an audio signal from our dataset with one from the MAESTRO dataset. From the analysis, we noticed many differences in sequences, noise, and tact expressed in waveforms. Another analysis was conducted on several melodic lines from our dataset and the MAESTRO dataset. Only a few examples are presented in this paper for comparison. These examples showcase the accuracy of note placement, chords, and tempo, with the three main dimensions of a melodic line being considered. This precise representation is a crucial step for future works and analyses. From the MAESTRO dataset, we have noticed many deficiencies and ambiguities in the sequential presentation where we have given the description and the reasons in that part, thus resulting in a very big difference in the beat as well.

The rest of the analysis was done with note-histogram, where we observed the collection of notes in both datasets for all melodic lines. In the beginning, the notes prove that our genre is unique, and the tonality and tempo are approximately the same in all the lines, resulting in that close and accurate collection, giving us an advantage in continuing to create other melodic sequences in this form. As for MAESTRO, we notice that the grades are very scattered and there is no determining grade such as in our dataset where B or A# has been decisive regarding the tonality in A minor of our sequences. That distribution clearly shows the lack of accuracy and the use of different tonalities, octaves, and beats, which not only change from one to another but also change within a melodic line, which is strictly prohibited according to music professionals, because when there is no foundation such as rhythm uniform, then the change in time of the sequences creates turbulence and disrupts the balance of a group of frequencies for which they give the meaning of musical continuity.

REFERENCES

1. Ali C. Gedik, B. B. (2010). Pitch-frequency histogram-based music information retrieval for Turkish music. *Elsevier*, 1049-1063. doi:10.1016/j.sigpro.2009.06.017
2. Anand, V. V. (2020). Music Genre Clasification with Deep Learning. *SCOPUS*, 1-6.
3. Briot, J. (2021). From artificial neural networks to deep learning for music generation: history, concepts and trends. *Neural Comput & Applic*, *Springer*, 31-65. doi:10.1007/s00521-020-05399-0
4. Chen, Y.-H. H.-Y.-H.-H. (2020). Automatic Composition of Guitar Tabs by Transformers and Groove Modeling. *Accepted at Proc. Int. Society for Music Information Retrieval Conf. 2020*, 1-5.
5. Chris Donahue, H. H. (2019). LakhNES: Improving multi-instrumental music generation with cross-domain pre-training. *arXiv:1907.04868 [cs.SD]*, 1-8.
6. Colton, B. B. (2021). A QUALITY-DIVERSITY-BASED EVALUATION STRATEGY FOR SYMBOLIC MUSIC GENERATION. *ML Evaluation Standards Workshop at ICLR*, 1-6.
7. Curtis Hawthorne, A. S.-Z. (2019). Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset. *In International Conference on Learning Representations, 2019.*, (p. Version 2.0).
8. DuBreuil, A. (2020). *Music Composition with Magenta.* Birmingham: Packt.
9. G´omez, N. K. (2015). Automatic Transcription of Flamenco Singing from. *arXiv:1510.04039v1 [cs.SD] IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, 1-16.
10. Gatti, G. O. (2017). Automatic generation of lyrics parodies. *In Proc. ACM Multimedia*, 485-491.
11. Hawthorne, I. S. (2021). Sequence-to-Sequence Piano Transcription with Transformers. *Sound (cs.SD); Machine Learning (cs:LG); Audio and Speech Procesing (eess.AS)*, 35-40.
12. I. Barbancho, L. J. (2012). Inharmonicity-Based Method for the Automatic Generation of Guitar Tablature. *IEEE Transactions on Audio, Speech, and Language Processing*, 1857-1868.
13. Kentaro Shibataa, E. N. (2021). Non-Local Musical Statistics as Guides for Audio-to-Score Piano Transcription. *arXiv:2008.12710v3 [cs.SD]*, 1-16.
14. Kong, Q. L. (2021). High-resolution piano transcription with pedals by regressing onsets and offsets times. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 3707–3717.
15. Martineau, J. (2008). The Elements of Music: Melody, Rhythm and Harmony. 1-6.
16. N. Rosmawarni, T. T. (2023). "Handcrafted Feature From Classification Mood Music Indonesia With Machine Learning BERT and Transformer. *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, 526-531. doi:10.1109/ICIMCIS60089.2023.10349006.

17. N. Rosmawarni, T. T. (2023). Hyperparameter Tuning On Machine Learning Transformers For Mood Classification In Indonesian Music. *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, 639-643. doi:10.1109/ICIMCIS60089.2023.10349008.

18. Nakamura, E. Y. (2017). Performance error detection and post-processing for fast and accurate symbolic music alignment. *In International Society for Music Information Retrieval (ISMIR)*, 347–353.

19. Oore, I. S. (2018). This time with feeling: Learning expressive musical performance. *Neural Computing and Aplications*, 50-62.

20. Peterson, G. H. (2017). Integrating Computer Science into Music Education. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 459-464.

21. Qiu, Z. Y. (2023). Generated Therapeutic Music Based on the ISO Principle. *Communications in Computer and Information Science, Springer,* 32-45.

22. Qiuqiang Kong, B. L. (2022). GiantMIDI-Piano: A large-scale MIDI Dataset for Classical Piano Music. *symarXiv:2010.07061v3*, 1-11.

23. Sarmento, P. K. (2023). GTR-CTRL: Instrument and Genre Conditioning for Guitar-Focused Music Generation with Transformers. *(eds) Artificial Intelligence in Music, Sound, Art and Design, Springer*, 260-275.

24. Shih-Lun Wu, Y.-H. Y. (2022). MuseMorphose: Full-Song and Fine-Grained Piano Music Style Transfer with One Transformer VAE. *arXiv:2105.04090 [cs.SD]*, 1-8.

25. Silla, A. L. (2017). Music education meets computer science and engineering education. *2016 IEE Frontiers in Education Conference*, 1-7.

26. Sreetama Mukherjee, M. M. (2022). ComposeInStyle: Music composition with and without Style Transfer. *ELSEVIER*.

27. Xianchao Wu, C. W. (2020). Transformer-XL Based Music Generation with Multiple Sequences of Time-valued Notes. *Multimedia (cs.MM); Audio and Speech Processing (eess.AS)*, 1-9.

28. Yu-Hua Chen, Y.-H. H.-Y.-H. (2020). Automatic Composition of Guitar Tabs by Transformers and Groove Modeling. *arXiv:2008.01431 [cs.SD]*, 1-8.