International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 74-85

https://doi.org/10.70517/ijhsa46406

Research on the Inheritance and Development of Yunnan Ethnic Minority Music Based on Computerized MIDI Sequence Editing Technology

Qingyuan Zeng[1,*]
[1] Yunnan College of Business Management, Kunming, Yunnan, 650300, China
Corresponding authors: (e-mail: 17279166990@163.com).

Abstract This paper proposes a MIDI automatic composition framework that integrates multi-track clustering algorithm and WaveNet model. The main melody is extracted by multi-track clustering algorithm, and the iterative prediction mechanism of pitch sequence is constructed based on WaveNet model. The model designed in this paper is used to generate Yunnan Ethnic Minority Music and explore its specific application effects. Deconstruct the mapping relationship between pitch and physical parameters, and quantify the short-time energy and spectral characteristics. The skyline algorithm is selected as a control to test the improvement effect of multi-track clustering algorithm in training efficiency and accuracy. Combined with user ratings and melody line visualization, the performance level of music generation of this paper's model is analyzed. The results show that the music generated by this paper's model improves about 24%~58% on five subjective evaluation indexes respectively compared to skyline model, and about 12%~22% on five subjective evaluation indexes respectively compared to KD3 model. This study provides a solution with both technical suitability and cultural fidelity for the digital inheritance of ethnic music, which is of great significance for the inheritance and development of Yunnan Ethnic Minority Music.

Index Terms MIDI, multi-track clustering algorithm, WaveNet model, Yunnan Ethnic Minority Music, music generation

I. Introduction

Each ethnic group has its own unique art and culture, and because of this, China's cultural resources are getting richer and richer. Among them, Yunnan ethnic Ethnic Minority Music has an important historical position that cannot be ignored in the traditional ethnic cultural resources, and is an indispensable and important part of the traditional ethnic cultural resources [1], [2]. The region's ethnic music is even unique artistic characteristics, and with the continuous integration of Western culture, Yunnan Ethnic Minority Music has been subjected to a certain impact and threat, and gradually appeared the phenomenon of dilution of national attributes, the literature [3] describes the great impact of modern civilization and foreign culture on the traditional art of the Yi ethnic group in Yunnan, so that it is facing the crisis of inheritance and analyzes its inheritance and development of effective measures and methods. Literature [4] reveals the great role played by ethnic music in protecting and perpetuating the complex cultural structure of the Bai ethnic group in Yunnan, ensuring its lasting resonance and cultural vitality from generation to generation, etc., and proposes measures for the protection and inheritance of music, including educational programs. Therefore, the inheritance and preservation of Yunnan Ethnic Minority Music has become a topic that needs to be focused on.

The traditional oral transmission system and industry occlusion have limited the inheritance and development of ethnic music, and in order to better protect and pass on ethnic music, music digital interface (MIDI) plays an important role [5]. Literature [6] examined the preparation of scores in MIDI format through mathematical modeling, revealing that the method can be used to analyze other types of abstract texts that are difficult to formalize in a variety of subject areas. In the production of ethnomusicology, all the steps can be done using a computer, and the parameters recorded by the MIDI communication protocol can be piggybacked on a software sound source in a music workstation to output the sounds of various instruments [7]-[9]. The convenience and diversity of this kind of production makes digital music flourish. Combining the characteristics of Yunnan ethnic music and the advantages of digital music production, MIDI sequence editor technology can realize the goal of automated generation and production of music with characteristics of ethnic music, promote the digital development and application of ethnic music, and make the ethnic music glow with new vitality and be better inherited and developed [10], [11].

Ethnic music integrates national culture, life and history, and is of great value to ethnic minorities and even to the entire Chinese nation. Literature [12] analyzes the historical development and dissemination of Miao music in Yunnan, emphasizes the significance of Miao music, and describes its ancient roots, its application in life, the impact of cultural exchanges it has had, and the important information it possesses. Literature [13] discusses the Dai music culture based on the literature review, and discusses it from many angles, and proposes strategies to cope with the problems existing in the protection and inheritance of the intangible cultural heritage of the Dai music culture. Literature [14] examines the challenges and opportunities of music transmission methods and the inheritance and development of folk music in the context of the digital era, and analyzes the impact of the transmission methods on the inheritance of folk music in order to continue to seek new ways of inheritance in order to promote the development of Chinese folk music. Literature [15] introduces the different types of Dong, points out the characteristics and shortcomings between them, as well as the types of their folk songs, and analyzes the stylistic characteristics of the two in terms of mood, tonality, and rhythmic aesthetics, in addition to considering the problems facing the development of Dong music at present. And with the development of information technology, its application in music has gradually been paid attention to, especially in music creation, inheritance and development, literature [16] constructed a model of ethnic music creation based on electronic music technology, and through the test of its function, it was pointed out that this creation mode greatly improved the accuracy of music creation, and it was more reasonable than the traditional processing model. Literature [17] describes the use of technology in music education and the positive impact it has had, especially in terms of increasing students' interest in music, but reveals that music technology is an afterthought and is not treated at a better and higher level in the curriculum of teacher training. Literature [18] emphasizes digital technology in music education and examines the current state of development of this emerging technology by discussing insights with a variety of stakeholders, including music educators, school administrators, and others. However, as of yet, no scholars have examined the heritage and development of ethnic Ethnic Minority Music from the perspective of computerized MIDI sequence editing technology.

Firstly, this paper describes the basic principles of music production in MIDI sequential editing technology by combining the characteristics of national music and the advantages of digital music production. Secondly, the clustering of track blocks on the basis of skyline algorithm realizes the extraction of main melody. Finally, iterative prediction based on WaveNet model is used to generate coherent ethnic style music segments. Taking Yunnan Ethnic Minority Music as the research object, 50 pieces of Yunnan Ethnic Minority Music are selected to construct the dataset. Combine short-time energy analysis and frequency domain processing to quantify the acoustic parameters. Set up a control test to verify the superiority of this paper's method by comparing the training efficiency and feature recognition accuracy. Using WaveNet model to generate chords, the feasibility of ethnic music stylization generation is assessed based on user evaluation. Based on the research results, the strategy of Yunnan Ethnic Minority Music inheritance and development is proposed.

II. MIDI-based modeling of automatic compositions

II. A.Fundamentals of Music Production in MIDI Sequence Editing Technology

With the rapid development of multimedia technology, the forms and applications of music creation have changed dramatically. In the era when digital music is very popular, many music producers take advantage of the convenience of digital music to produce music, using the Music Digital Interface (MIDI) communication protocol to record the digital parameters required by the music for arranging the music, which is much more convenient than relying on the microphone to record the performance of a physical instrument. In the process of digital music production, all steps can be done using a computer, and the parameters recorded by the MIDI communication protocol can be equipped with a software sound source in a music workstation (DAW) to output the sounds of various instruments. The convenience and diversity of this kind of production makes digital music flourish.

Combining the characteristics of ethnic music and the advantages of digital music production, the paper uses the supervised learning model of machine learning to train the digital parameters recorded in the MIDI sequence editor to achieve the goal of automated generation and production of music with ethnic music characteristics, to promote the digital development and application of ethnic music, and to make ethnic music revitalized. In order to cope with the diversified demands of the market for music genres and the uniqueness of folk music, the experiments in this paper will train the MIDI data of many different folk music styles. By training a perfect automated composition model of ethnic music, a large number of ethnic music styles can be provided to the music market, and time and labor costs can be saved, creating more directions for the application and development of ethnic music.

Music is a form of artistic expression composed of sound, and the main constituent elements of music are pitch, rhythm, and timbre. In different music styles, different elements are emphasized to different degrees. Pitch is an important feature of folk music, representing the frequency of sound. For stringed instruments in folk music

performance, there are three ways to change the pitch, which are to adjust the length, tension, and density of the strings.

(1) The longer the string length, the lower the pitch; the shorter the string length, the higher the pitch. The frequency of vibration is inversely proportional to the length:

$$f \propto \frac{1}{l} \tag{1}$$

(2) The lower the tension on the string, the lower the pitch; the higher the tension, the higher the pitch. The frequency of vibration is proportional to the square root of the tension:

$$f \propto \sqrt{T} \tag{2}$$

(3) The higher the density of the string, the lower the pitch; the lower the density, the higher the pitch. The frequency of vibration is inversely proportional to the square root of density:

$$f \propto \frac{1}{\sqrt{\rho}} \tag{3}$$

where, f denotes vibrational frequency, l denotes length, T denotes tension, and ρ denotes density.

Tone color is another important characteristic of folk music. The main factor contributing to differences in timbre is the difference in the composition of overtones produced by the vibration of sound waves. The main factors that differentiate human perception of sound are waveform, sound pressure and spectrum.

Waveform is the shape of the sound wave and is the largest contributor to differences in timbre. Waveform is controlled by 4 parameters including:

(1) Attack (A): the time it takes for the sound to go from nothing to peak;

(2) Decay (D): the time it takes for the sound to go from peak to a smooth state;

(3) Sustain (S): the time the sound is in a smooth state;

(4) Relcase (R): the time when the sound decays from a smooth state to nothing.

The music waveform trend graph is shown in Figure 1.

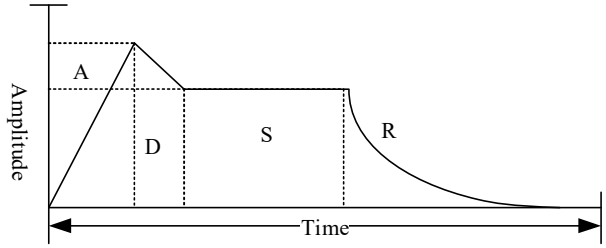


Figure 1: Music waveform trends

Sound pressure is the sound wave propagation in the medium, due to the vibration of the pressure variable, the symbol is p, the unit is Pa, often use the sound pressure level SPL to express the size of the sound pressure. Sound waves transmitted in the medium, the density of the media particles with the sound wave changes, the instantaneous sound pressure at each point will be different. Therefore, taking the root mean square $p_{\max}$ as its average value to calculate, the relationship between the peak value of the sine wave generated by the sound p_{park} and the root mean square $p_{\max}$ is as follows:

$$p_{\max} = \frac{p_{park}}{\sqrt{2}} \tag{4}$$

In ethnomusicological compositions, the stimulation of the human senses by the sound pressure level often lies in matching the rise and fall of the song. In music, the sound spectrum represents the frequency performance of a sound in a temporal sequence, and it is common to use time-frequency data to characterize the change of the sound spectrum with the temporal sequence.

II. B.Multi-track clustering main melody extraction algorithm

The multi-track clustering algorithm is based on the skyline algorithm, which realizes the extraction of the main melody by clustering the track blocks.

The steps of multi-track clustering algorithm are as follows:

(1) Generate the note matrix

(2) Use skyline algorithm

The skyline algorithm is executed for each track block with the purpose of ensuring that each track block contains only a single melody.

(3) Find and weight the average pitch distribution vector

There are 12 notes in an octave in the piano, each representing a different pitch, so we define 12 pitch values within a piece of music. Equations (5) and (6) are used to find the average pitch distribution vector a for the entire piece of music as well as the number of times a note with a pitch of p occurs a_p in the N track blocks.

$$a = \left(\overline{a_1}, \overline{a_2} \ldots \overline{a_{12}}\right) \tag{5}$$

$$a_p = \frac{\sum_{K-1}^{N} a_{kp}}{N} \tag{6}$$

Equation (7) is used to find the weighted average pitch distribution vector $\overline{a_w}$ for the whole music, and Equation (8) is used to find the number of times a note with pitch p occurs a_{wp} in a block of N tracks, where d_k is the ratio of the number of notes in the k th track to the number of all notes in the whole music.

$$\overline{a_w} = \left(\overline{a_{w1}}, \overline{a_{w2}} \ldots \overline{a_{w12}}\right) \tag{7}$$

$$a_{wp} = \sum_{K=1}^{N} a_{kp} d_k \tag{8}$$

(4) Clustering the mean pitch distribution vector

Before clustering the mean pitch distribution vectors, the distance $d\left(a_m, a_n\right)$ between the two track blocks a_m and a_n needs to be derived from equation (9).

$$d\left(a_m, a_n\right) = \sqrt{\sum_{\kappa=1}^{12} \left(a_{mk} - \overline{a_{nk}}\right)^2} \tag{9}$$

It is also necessary to fix a threshold distance as a reference standard, as shown in Equation (10).

$$y = \frac{d\left(\overline{a_w}, \overline{a}\right)}{2} \tag{10}$$

The vector distribution of each track block is an independent cluster, the distance between every two clusters is calculated, the two closest clusters are found, if the distance is less than the threshold distance, the two clusters are merged and the calculation of the distance between every two clusters is continued; if the distance is greater than the inter-value distance, the calculation is stopped. In this way, multiple merged clusters are obtained.

(5) Pick the block of tracks with the highest combined significance in each of the clusters

The information entropy $h(x)$, $f(p)$ of each track block is found using equation (11), and 2 is the frequency of occurrence of notes of pitch p in each track block, where the base of the $\log$ function is 12, with the aim of having $h(x)$ taken in the range 0-1.

$$h(x) = -\sum_{p=1}^{12} f(p) \log_{12} f(p) \tag{11}$$

Then from the clusters obtained after clustering in the third step, the track block that can represent the cluster, i.e., the track block with the highest combined significance, is picked. The combined significance z_k of each track block is given by Equation (12), and $\overline{f_k}$ is the average pitch in each track.

$$z_k = \overline{f_k} + 128 \times h(x) \tag{12}$$

(6) The main theme is extracted using the skyline algorithm again

Setting all the notes of the track block representing the community, after executing the skyline algorithm on these notes, the main melody in the whole music is extracted.

II. C.Automatic Composition Modeling

For automatic composition, the WaveNet model was chosen as the training model for this experiment, which is a sequence generation model, and can also be described as a causal convolutional model, since its essence is to make predictions of possible future outcomes through convolutional computation, and then use the predicted values as part of the inputs to make subsequent predictions, and keep looping in order to achieve the purpose of automatic composition. In this subsection, the principles and structure of the model implementation are described.

II. C. 1) Principles of automatic composition

After the feature extraction of the MIDI files, we obtained the monophonic and chordal sequences of all the instruments in the dataset. The principle of automatic composition used in this project is similar to that of typing, in which the input method suggests the upcoming words through the existing text. For this project, an array of monophonic and chordal sequences of length 32 is chosen as the input to the model, the next monophonic or chordal sequence that is about to appear is predicted, and the prediction result is taken as the output, and then the output is taken as a part of the array of pitch sequences as the input to continue the prediction, and in this way a sequence of prediction results is obtained, which is the result of the pitch data of the composition, and the principle of the project is shown in Fig. 2. The principle is shown in Fig. 2.

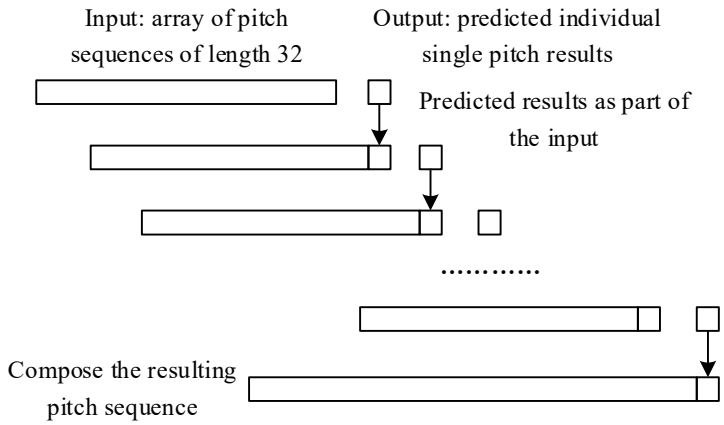


Figure 2: Principle of automatic composition

II. C. 2) WaveNet model structure

The basic principle of WaveNet is to predict the value of the t nd point based on the first $t-1$ point of the sequence, so when we know the sequence of notes in a piece of music, we can use the model to predict the subsequent possible notes, the basic formula for WavcNct prediction is shown in equation (13).

$$p(x)=\prod_{t=1}^{T} p\left(x_t \mid x_1, x_2, \ldots, x_{t-1}\right) \tag{13}$$

The WaveNet model is a convolutional model with a multi-layer convolutional network structure, where each convolutional layer performs a convolutional operation on the data from the previous layer, and the larger the convolutional kernel is during the operation, the larger the perceptual range of the layer is, and the better the perceptual ability in the time domain is. When the output layer gets the final prediction at the end of the convolutional layer's operation, it outputs the prediction and uses that result as part of the input for subsequent iterations in a continuous loop.

III. Example analysis of MIDI-based generation of Yunnan Ethnic Minority Music

This chapter focuses on the analysis and processing of a library consisting of 50 Yunnan Ethnic Minority Music tracks collected from the Internet, with musical characteristics such as pitch beat, volume, track, and onset obtained through MIDI.

III. A. Extraction of main theme features

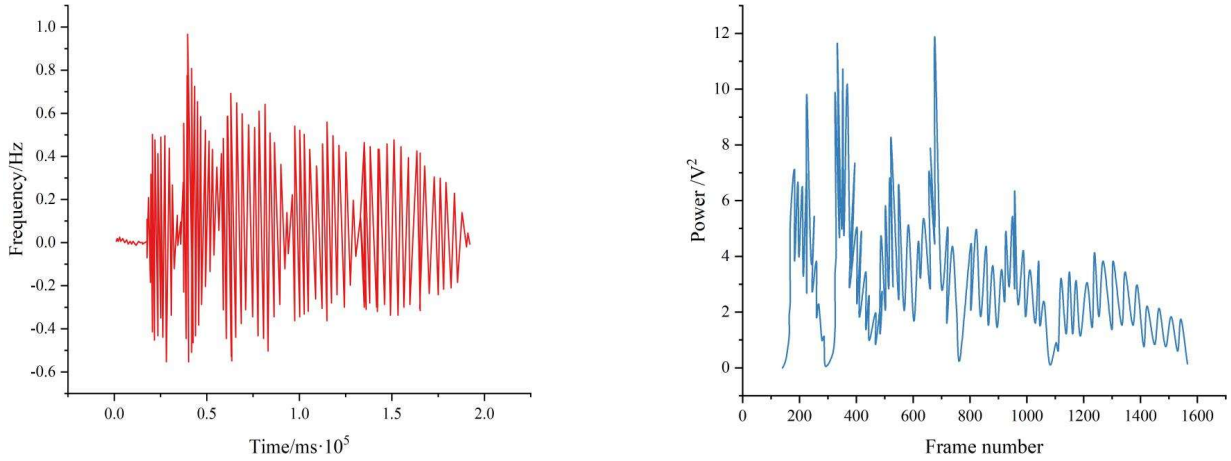
III. A. 1) Short-term energy analysis

The energy of an audio signal varies relatively significantly with time, and its short-time energy analysis gives a suitable description of the response to these amplitude changes. For signal $\{x(n)\}$, the short-time energy is defined as follows:

$$E_n=\sum_{m=-\infty}^{\infty}[x(m)w(n-m)]^2=\sum_{m=-\infty}^{\infty}x^2(m)h(n-m)=x^2(n)*h(n) \tag{14}$$

where $h(n)=w^2(n)$. Eq. (14) represents the short-time energy when the window function is added starting at the n nd point of the signal. It can be seen that the short-time energy can be viewed as the square of the audio signal passing through the output of a linear filter that has a unit impulse response of $h(n)$.

The short-time energy can effectively determine the magnitude of the signal amplitude and can be used to make a sound/no sound determination. For Yunnan Ethnic Minority Music processing, the sound signal is shown in Fig. 3(a), and the output signal after short-time energy analysis is shown in Fig. 3(b). 0~100 frames of short-time energy is close to 0, which may be a silent section or background noise; the energy rises significantly from 100 to 800 frames, which corresponds to the main body of the music; and the energy decreases significantly after 800 frames, which corresponds to the decay stage of the music.



(a)Sound signal (b)Short-time energy

Figure 3: The sound signal passes through short-time energy

III. A. 2) Frequency domain processing

The sound signal is further processed and its power spectrum is obtained by self-multiplication after Fourier transform, and according to the nature of Fourier transform, the power spectrum is symmetrical between left and right. Since the processed data object is discrete, the result is also a series of discrete quantities, the number of these discrete points is the number of sampling points contained in each frame, so in the feature extraction, for the power spectrum only consider the left half of the points, i.e., the number of discrete points considered is half of the length of the frame, and the left portion of the power spectrum is shown in Figure 4.

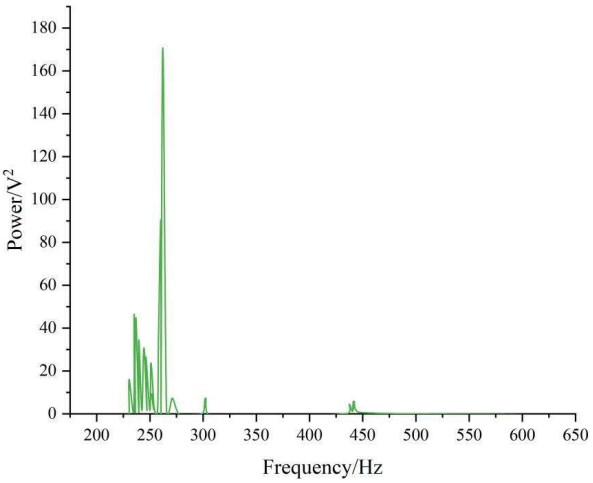


Figure 4: Left part of the power spectrum

The audio stream in this paper has a sampling rate of 11050 HZ, and the highest frequency that can be recorded is half the sampling rate. These discrete points correspond to the frequencies uniformly in proportion. The fundamental frequency of a set of harmonics is required, i.e., the frequency corresponding to the discrete amount of the leftmost crest of that harmonic is found. For example, the frame length is 550, the sampling rate of the audio stream is 11050 HZ, and the point corresponding to the fundamental frequency peak in the group of harmonics is

the 23rd point from left to right. The frequency of the set of harmonics is then: $23*[(11050/2)/(550/2)]=462HZ$, and its error is: $[(11050/2)/(550/2)]/2HZ$ about 10 HZ. Reducing the fundamental frequency extraction error needs to start from the process of fundamental frequency extraction. Let the sampling rate be T, the frame length of the sub-frame processing be $frame\_len$, and the number of harmonics be n. Then the formula for the frequency error is:

$$f_e=\frac{1}{2}*\frac{1}{n}*\frac{T/2}{frame\_len/2}=\frac{T}{2*n*frame\_len} \tag{15}$$

From equation (15), it can be seen that to reduce the error rate f_e, the sampling accuracy of the audio stream can be reduced or the length of the frame can be increased. If the sampling accuracy is reduced, some high frequency harmonics will be lost, which will make the confirmation of the fundamental frequency difficult, and the reduction in the number of harmonics, i.e., n in Eq. (15) becomes smaller, which is also not conducive to reducing the frequency error. Reducing the sampling rate will also lead to a relatively long frame length. Increasing the frame length arbitrarily is also not allowed, the sound is only smooth for a short period of time. If the frame length is made too long, it may superimpose multiple sounds of different frequencies at different moments in time, and it is impossible to distinguish the sequence of these different sounds due to the loss of time information in the FFT variation. The sound is generally stable in 30ms-50ms. Let the time length of each frame be $frame\_t$ then there is:

$$frame\_t=\frac{frame\_len}{T} \tag{16}$$

Substituting equation (16) into equation (15) has:

$$f_e=\frac{1}{2*n*frame\_t} \tag{17}$$

When the fundamental frequency is relatively low, the number of harmonics n will be relatively large. When the fundamental frequency is higher, the number of harmonics n may be relatively small, but the frequency difference between the higher frequency semitones is also larger. Therefore take plus $frame\_e=0.5m$, which can basically satisfy the need of fundamental frequency extraction. Therefore the frame length $frame\_len=550$ is chosen in this paper.

In the selected Yunnan Ethnic Minority Music, there may be background noise, current noise, and the music in the music library is mixed with human voices and various instruments, etc. Therefore, the spectrograms of some frames may be different. Therefore, the spectrograms of some of the frames are particularly complex, and the processing method for these complex frames is to separate the harmonics of each group, and take the tone with larger energy in the middle as the main melody of the music: frames in which the harmonics cannot be detected are treated as blank tones.

Let $fft(n)$ denote the power spectrum sequence after FFT transform (Fast Fourier Transform) and after convolution and logarithm, since only the first half of the sequence is taken into account, there are $1\le n\le\frac{frame\_len}{2}$ and n integers, and $frame\_len$ denotes the frame length. The extraction of the fundamental frequency is divided into only two steps: determining the discrete quantities corresponding to the wave peaks in the power spectrum: dividing these wave peaks into groups of harmonics, and taking the group of harmonics with the highest power to be identified as the main schlieren.

The algorithm for determining the wave peaks in the power spectrum is equation (18), i.e., only the wave peaks in $fft(n)$ are retained, and other values are set to zero, and $FFT(n)$ is used to denote the array after extracting the wave peaks, and $1\le n\le\frac{frame\_len}{2}$, n are integers. Since there may be a large number of wave peaks in $fft(n)$ that are caused by disturbances such as noise, a portion of the wave peaks with smaller peaks are discarded.

Among them:

$$FFT(n)=\begin{cases} fft(n-1)\le fft(n)\le fft(n+1) \\ fft(n) \quad And\ fft(n)>\alpha\overline{fft(n)},1<n\le\frac{frume\_len}{2} \\ 0 \qquad\qquad Other \end{cases} \tag{18}$$

$$\overline{fft(n)} = \frac{\sum_{n=1}^{\left\lfloor \frac{frume\_len}{2} \right\rfloor} fft(n)}{\left\lfloor \frac{frume\_len}{2} \right\rfloor} \tag{19}$$

α is an empirical parameter indicating the minimum closure value of the wave crest, which takes the value of $\alpha = 3$ in this system.

Equation (20) can be used to find the fundamental frequency. Where T denotes the sampling rate and $frame\_len$ denotes the frame length.

$$f_{nase} = position * \frac{T}{frame\_len} \tag{20}$$

In order to discretize the human auditory frequency range into a number of semitones, it is therefore also necessary to convert the fundamental frequency to be represented by a semitone, which can be done by using $Semitone = 12 \times \log_2 (freq / 440) + 69$. Finally, all fundamental frequencies are converted into semitone units.

III. A. 3) Extraction of results

Since the main theme extraction method in this paper is an improvement on the skyline algorithm, the experimental results are also compared with the skyline algorithm. First of all, in the training efficiency, the training time comparison results are shown in Figure 5. The red part of it is the iteration error of this paper's method with the number of iterations, while the blue part is the comparison method. The error of this paper's method decreases rapidly in about 25 rounds, and then it tends to level off, and the control method converges more slowly, and it can only reach a lower error in about 50 rounds, and the final error value is still larger than that of this paper's method, from which we can see that this paper's method has a certain degree of improvement in the training efficiency.

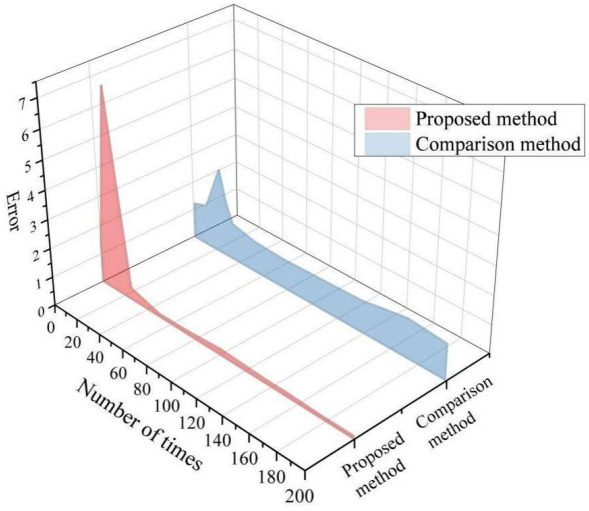


Figure 5: Comparison of training time

Next the recognition rates of the two methods were compared. By analyzing the MIDI file format, the note information of each MIDI can be obtained. It was verified by CakeWalk that the note information extraction part was correct. We compared the recognition rates from two different perspectives. The first way is to randomly select half of all labeled MIDI files as a training set and the other half as a test set. The feature vectors of each track can be calculated from the note information of the MIDI. The feature vectors of the training set are used as input to train the two classification models separately. At the end of training test the accuracy of the classification models with the test set's and record the test results. Repeat the above process 10 times.

The accuracy of the two methods is shown in Fig. 6. From the figure it can be seen that the average recognition accuracy of audio track features based on this paper's method is 82.8%, while the average recognition accuracy of the comparison method is 71.4%.

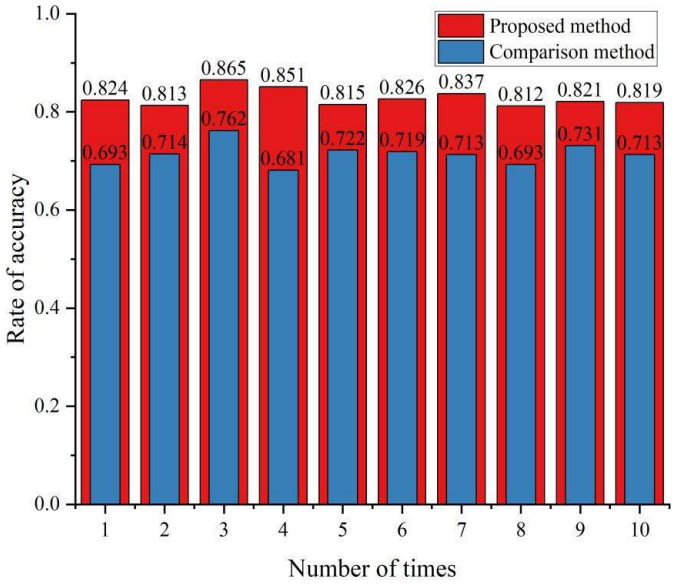


Figure 6: Comparison ofrecognition accuracy

The final extraction results are shown in Table 1. As can be seen from Table 1, the track discrimination values of track 4, track 11, and track 2 are 0.042, -0.021, and -0.038, respectively, so track 4 is the main melody track.

Table 1: Main melody feature extraction results

Sound track number	Track 4	Track 11	Track 2
Articulation time	00:45:200	00:45:200	00:45:400
Voice volume	0.7937	1	1
Right and left channel balance	1	0.4	0.6
Average strength	0.9949	0.7051	1
Interval ratio	0.8	1	0.9936
Mean length	72.4867	78.3982	78.3982
Track discrimination value	0.042	-0.021	-0.038

III. B. Automatic chord generation

After determining the tonality of the melody, the chords are further selected to fit the tonality of the main melody. A chord is composed of three or more tones. A triad is a chord in which the three tones are superimposed in thirds, and its structure consists of a root, a third, and a fifth, the third being in thirds from the root, and the fifth being in fifths from the root. The root method refers to a method used in music theory to represent chords by naming the notes of the chord according to their position on the scale. In a chord, the most fundamental note is called the “root note”, which determines the name and character of the chord.

In this paper, we use the WaveNet model for chord generation, the input of the model is a 32-long array of single notes and chord sequences, which is first convolved by a multilayer convolutional layer, and then predicted by a fully connected layer, and the predicted chord results are the output of the model. The MIDI pitch representation of the output result is shown in Table 2, with 58 as the center C. From Table 2, we can obtain: I ={45,50,54,58,62,66,71,74,78}, II ={52,55,59,64,67,71,76,79,83}, III={51,55,59,63,68,72,75,81,85}, and similarly, the Other chord progressions can be represented by the corresponding sets.

Finally, the generated chords are shifted up or down an octave according to the pitch of the melodic dominant to realize the generation of Yunnan Ethnic Minority Music.

III. C. Experimental evaluation

In order to verify the feasibility of the experimental model of this paper in generating Yunnan Ethnic Minority Music, a comparative evaluation of the generated results is needed, and here skyline model, KD3 model and the model of this paper are used to do the comparative analysis. In order to ensure the fairness of the evaluation, 10 pieces of music were generated by each of the three methods, all of which used the collected 50 Yunnan Ethnic Minority

Music datasets, and the same parameter control was chosen to evaluate the 10 randomly generated pieces of music.

Table 2: MIDI pitch representation of output results

Series	Chord	Root	Tritone	Five tone
Ⅰ	C	45	50	54
		58	62	66
		71	74	78
Ⅱ	Dm	52	55	59
		64	67	71
		76	79	83
Ⅲ	Em	51	55	59
		63	68	72
		75	81	85
Ⅳ	F	52	56	58
		64	69	72
		76	82	86
Ⅴ	G	54	58	60
		67	70	72
		80	82	84
Ⅵ	Am	58	62	68
		70	74	79
		82	86	90
Ⅶ	Bdim	61	64	67
		73	76	79
		85	88	91

The evaluation of the music piece is an aesthetic activity, and the aesthetic activity is jointly accomplished by the subject of appreciation and the object of the music piece. In order to ensure the accuracy of the evaluation, the selection of the subject and object needs to have certain conditions, the subject needs to have aesthetic skills and be able to feel the music, and the object needs to have an aesthetic foundation and its own sense of beauty. Subjective evaluation of the aesthetics of musical works is a shallow to deep process, from intuitive experience to emotional expression to resonance, for these three levels, using the five indicators often used to evaluate music, respectively, for the music of the ethnicity, structure, pleasant to the ear, the degree of association, resonance. Since aesthetic differences exist objectively, 25 students each from music majors and non-music majors were invited to give ratings to the generated scores in each of these 5 aspects. The score of subjective evaluation was set at 1-10 points, and the results of the subjective evaluation of the composition examples are shown in Table 3, where the values are averages.

Table 3: Comparison of subjective evaluation results

Subjective evaluation index	skyline		KD3		Proposed	
	Music major	Non-music major	Music major	Non-music major	Music major	Non-music major
National character	7.13	7.35	6.88	7.92	8.76	9.15
Structural	7.05	7.23	8.28	8.44	9.72	9.75
Euphonicity	5.23	6.08	7.19	7.28	8.82	8.82
Association degree	5.33	6.42	6.48	7.03	7.28	7.84
Resonance	5.07	5.39	6.29	6.97	7.47	7.59

Comprehensive evaluation value = the average value of the evaluation of music majors * 0.6 + the average value of the evaluation of non-music majors * 0.4, the subjective comprehensive evaluation results of the five indicators of the SKYLINE model are 7.22, 7.12, 5.57, 5.77, 5.20, respectively, and the subjective comprehensive evaluation results of the five indicators of the KD3 model are 7.30, 8.34, 7.23, 6.70, respectively, 6.56, and the subjective comprehensive evaluation results of the five indicators of this paper's model are 8.92, 9.73, 8.82, 7.50, 7.52. This

paper's model improves about 24% to 58% on the five subjective evaluation indicators respectively compared to the skyline model, and improves about 12% to 22% on the five subjective evaluation indicators respectively compared to the KD3 model.

The melodic lines of the music generated by this paper's model, the KD3 model, and the skyline model are shown in Figure 7. The rhythm of the melodic phrases generated by this paper's model is square, and most of the phrases in each melody end with a short rhythm into a long rhythm. On the other hand, the rhythms of the melodies generated by the skyline model and the KD3 model are more scattered, which makes it difficult to have a clear grasp of the length of the phrases, and also results in the melodies not having a relatively stable sense of suspension. Overall, the model in this paper has some advantages in the generation of Yunnan Ethnic Minority Music.

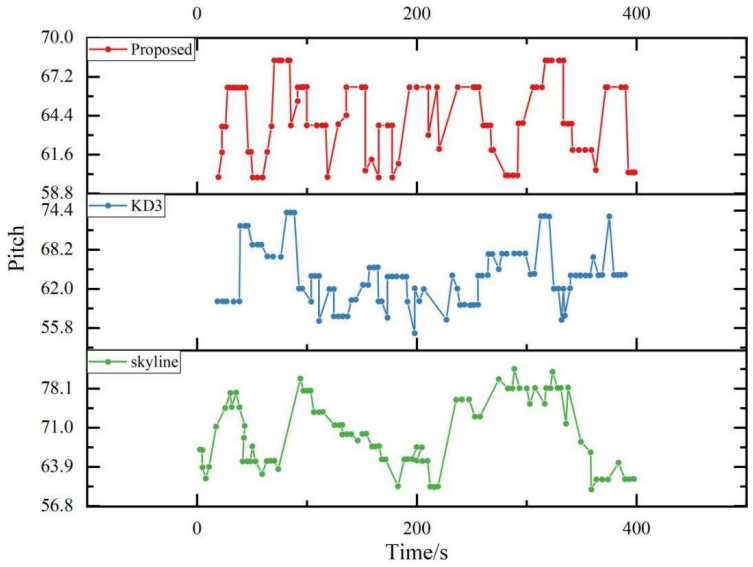


Figure 7: The melodic lines of the music generated by the model

IV. Conclusions and strategies

IV. A. Conclusion

In this paper, we propose an automatic composition model based on MIDI feature analysis and deep learning, and launch the application analysis with Yunnan Ethnic Minority Music as the research object.

In terms of training efficiency, the error of this paper's method decreases rapidly in about 25 rounds, and then tends to level off. The control method converges more slowly and can only reach a lower error in about 50 rounds, and the final error value is still larger than that of this paper's method, from which it can be seen that this paper's method has a certain improvement in training efficiency. In terms of feature recognition rate, the average recognition accuracy of the track features based on this paper's method is 82.8%, while the average recognition accuracy of the comparison method is only 71.4%.

Using WaveNet model for chord generation, the output Ⅰ ={45,50,54,58,62,66,71,74,78}, Ⅱ ={52,55,59,64,67,71,76,79,83}, Ⅲ ={51,55,59,63, 68,72,75,81,85}, and other chord progressions can be represented by the corresponding sets The other chord levels can also be represented by the corresponding sets. The generated chords are shifted up or down an octave according to the pitch of the melodic dominant to realize the generation of Yunnan Ethnic Minority Music. The music generated by the model in this paper is improved by about 24%~58% in five subjective evaluation indexes compared with the skyline model, and improved by about 12%~22% in five subjective evaluation indexes compared with the KD3 model. The rhythms of the melodic phrases generated by this paper's model are square, while the rhythms of the melodies generated by the skyline model and the KD3 model are more scattered, which proves that this paper's model has an advantage in generating music of Yunnan ethnic minorities.

IV. B. Strategies

This study provides an efficient and scalable technical path for the digital protection of ethnic music and the generation of multi-style music, for this reason, this paper puts forward a three-point strategy for the inheritance and development of ethnic Ethnic Minority Music in Yunnan.

(1) Construction of Living Inheritance System

Modeling and preserving the endangered musical varieties of Yunnan Ethnic Minority Music, including acoustics, performance method and cultural context. Establish a digital database containing Yunnan minority traditional music, instrumental music, songs and dances, and provide convenient search and analysis tools. Generate music that meets the characteristics of Yunnan's ethnic minorities through MIDI sequence editing technology. Design a cultural adaptation assessment model to quantitatively assess the cultural fidelity of innovative works.

(2) Cross-cultural integration and innovation

Utilize MIDI sequencing editing technology to integrate and innovate with Yunnan's traditional ethnic music, integrate elements from different ethnic groups in modern music creation, and form a new type of music style with Yunnan's local characteristics. Supporting cross-border cooperation and innovation between Yunnan traditional ethnic music and modern popular music, electronic music and other forms of music can not only preserve the unique charm of Yunnan Ethnic Minority Music, but also attract the interest of the younger generation.

(3) Establishing Protective Policies

The government should increase support for minority cultures, especially Ethnic Minority Music, and provide special funds to support programs and cultural protection policies to encourage cultural inheritance and innovation. Strengthen the protection of intellectual property rights for ethnic Ethnic Minority Music to ensure the originality and reasonable earnings of traditional music, and avoid the theft and improper commercialization of music and cultural resources.

References

[1] Kuang, J., & He, L. (2022). From oblivion to reappearance: A multi-faceted evaluation of the sustainability of folk music in Yunnan province of China. Sage Open, 12(3), 21582440221117806.

[2] Yuxin, Z., & Hirunrux, S. (2022). China's Cultural Policies and Countermeasures for the Protection and Development of Ethnic Music Education in Yunnan. Journal of Modern Learning Development, 7(10), 364-373.

[3] Zhipeng, D., Maneewattana, C., & Xiulei, R. (2024). The Definition of the Concept of Yunnan “New Folk Songs”. Journal of Roi Kaensarn Academi, 9(11), 651-664.

[4] Jiayang Li, D. F. A., & Su, Y. (2024). Exploring the Significance of Traditional Music in Safeguarding and Transmitting Intangible Cultural Heritage: A Case Study of the Yunnan Bai Ethnic Group. Cultura: International Journal of Philosophy of Culture and Axiology, 21(3).

[5] Wang, L., & Thoard, M. (2022). Research on the Inheritance and Development of Folk Music Among Primary and Secondary School Students Take Primary and secondary Schools in Hunan Province as Examples. Journal of the Association of Researchers, 27(4), 210-244.

[6] Gorbunova, I. B., & Chibirev, S. V. (2019). Modeling the process of musical creativity in musical instrument digital interface format. Opción: Revista De Ciencias Humanas Y Sociales, (22), 392-409.

[7] Kapoyos, R. J., Suharto, S., & Syakir, S. (2022). Bia Music: Traditional Music Heritage and Preserving Tradition Across Generations. Harmonia: Journal of Arts Research and Education, 22(2), 298-310.

[8] Mores, R. (2018). Music studio technology. Springer Handbook of Systematic Musicology, 221-258.

[9] Lehrman, P. D., & Tully, T. (2017). What is MIDI?. ed: Medford, MA: MMA.

[10] Mermikides, M. (2023). Remapping and relearning the guitar's pitch matrix with MIDI and Max/MSP. 21st Century Guitar: Evolutions and Augmentations, 17.

[11] Jones, D. (2016). The Complete Guide to Music Technology. Lulu. com.

[12] Yu, L., & Choatchamrat, S. (2024). Historical Development of Education and Learning in the Transmission of Miao Nationality Music in Yunnan Province, China. Journal of Education and Learning, 13(3), 113-122.

[13] Zhong, W. (2017, June). Reflections on the Development and Inheritance of the Dai Music Culture. In 2nd International Conference on Contemporary Education, Social Sciences and Humanities (ICCESSH 2017) (pp. 563-565). Atlantis Press.

[14] Xu, Y. (2023). Exploration of the Influence of Music Communication Methods on the Inheritance of Ethnic Minority Music. Art and Performance Letters, 4(8), 44-49.

[15] Xia, D. A. N. (2015). On style features, inheritance and development of folk songs in south and north dong minority. Higher Education of Social Science, 9(1), 48-51.

[16] Zhang, J. (2024). Research on the Application of Computer-Aided Electronic Music Technology in Folk Music Creation. International Journal of High Speed Electronics and Systems, 2440032.

[17] Dorfman, J. (2022). Theory and practice of technology-based music instruction. Oxford University Press.

[18] Thompson, D. (2022). Music Education Technology Curriculum and Development in the United States: Theory, Design, and Orientations. Kent State University.