Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on Multimodal Generation Strategy of Opera Style and Music Melody Based on Time Series Analysis

Jiping Liu[1], Mei Huang[1,✉]

[1] *Art College, Wanxi College, Lu'an, Anhui, 237012, China*

ABSTRACT

The application of artificial intelligence on the field of art can be used to assist the creation of musicians and provide new creative ideas for musicians. In this paper, firstly, an ARIMA model is established for the prediction problem of opera style, which is used to predict the trend of the development of opera style sequence, and the best model is selected according to the minimum information criterion and Bayesian criterion. Then an automatic music melody generation method based on the generative adversarial network framework is proposed, which applies the trained natural language generation model to music generation to textualize the music melody and reduce the model running time. In addition to this a barization music melody generation method is also used, which divides a large music melody into melodic segments and generates them segment by segment, reducing the difficulty of the model in generating the music melody. Finally, the Fourier transform method is used to extract the features of the music melody and complete the visualization of the music melody. The model ARIMA(2,1,1)(2,1,0)12 that best fits with the time-series prediction of the development of opera styles was identified through empirical analysis. The PB value of Leak-GAN_2 model in this paper is improved by 41.38% compared with MusicGAN. It shows that both the opera style prediction model and the music melody multimodal generation model constructed in this paper have better effect and certain advancement.

Keywords: Natural Language Processing, Generative Adversarial Network, ARIMA Model, Fourier Transform, Music Melody Multimodal Generation

1. Introduction

Opera is a unique and charming part of Chinese traditional cultural treasury, which not only carries ancient history and culture, but also is a unique performing art form, and with the emphasis on traditional culture and globalization, it is increasingly in demand and gradually has an irreplaceable market share [1-3]. Opera has unique artistic styles and expressions in different regions, periods and singers, such as the high-pitched voice of Peking Opera, the distinctive and exaggerated voice of

✉ Corresponding author.
E-mail address: hmcd--1999@163.com (M. Huang).
Received 05 March 2024; Revised 10 May 2024; Accepted 05 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-321

Sichuan Opera, and the dynamic and graceful voice of Cantonese Opera [4-7]. Through the changes of pitch, volume and timbre, the actors are able to express the character and emotion of different characters, showing the unique style of opera music [8]. Whether it is the high-pitched and impassioned recitation or the delicate and soft singing, the audience can deeply feel the inner world of the characters.

In the melody of opera music, singing is the main part of opera music, which is the main means to express the thoughts and feelings of the characters, delineate the character, and is also the main factor to determine the stylistic characteristics of a play [9]. The singing form of opera is mainly solo, used to express the character's thoughts and feelings and personality, there are also a small number of duets and sing-alongs, duets are used for repartee, and sing-alongs are used for mass scenes; the group singing of high-cavity operas with the help of the cavity, and the solo singing of the actors in the tone, volume and musical character of the formation of a stark contrast, which has a special artistic effect [10-11]. In addition, some of the opera close to the natural language or with rap singing, mostly used for narrative; melodic, far from the natural form of the language singing, mostly used for lyrical; melodic ups and downs, rhythm and speed changes in the singing, mostly used for strong emotional changes and dramatic conflicts in the scene [12-13]. Long sections of the cantata, often due to the lyrics of the plate or the song plate of the more different, both narrative and lyrical nature. Therefore, the need, in order to better study, opera.

Multimodal generation, on the other hand, converts and generates data in different modalities, and is commonly used in the fields of natural language processing, computer vision and speech synthesis. The goal of multimodal generation is to convert one modal data to another modality, or generate another modal data according to one modal data [14]. Currently, the study, promotion, and inheritance of opera are facing the challenge of decline, and there is a need to understand more about the style of opera and the characteristics of the meter in order to better carry on the inheritance [15]. Therefore, the study of multimodal generation of opera style and music melody is of great significance for the inheritance of opera.

In this paper, an opera style prediction model based on ARIMA model is first designed to collect the opera development style data from the Chinese Opera Research Center from January 2008 to December 2021 to estimate the future style development trend of opera. The ARIMA model that best fits the research direction of this paper was searched on the established training set. Based on this, a Leak-GAN model based on barization generation is investigated for the problem that most of the music melodies are too long. In order to visualize the results of music melody generation, the Fourier transform method is applied to extract the music features. Finally, SPB and QN metrics are then used to compare this paper's model with other models to verify the effectiveness of this paper's method.

2. Prediction of Opera Style Based on ARIMA Modeling

ARIMA model [16], also known as differential moving average autoregressive model, belongs to the time series analysis method, ARIMA consists of differential operations and autoregressive moving average model (ARMA model).ARMA model contains two parts, namely autoregressive model (AR model) and moving average model (MA model), the combination of the AR model and the MA model constructs the ARMA model. Set a set of random variables in chronological order.

2.1. ARIMA model construction idea

Smoothing is when the data fluctuates up and down around some constant with a limited range of fluctuations, when there is a constant mean and a constant variance, and smoothing is categorized into strict and wide smoothing.

Strictly smooth is the joint distribution of the variables at all points in time does not change with time translation, i.e., the time series $\{Y_t, t \in T\}$ is strictly smooth if the joint distribution of the random

variable $Y_{t_1}, Y_{t_2}, \cdots, Y_{t_t}$ for moment $t_1, t_2, \cdots, t_l$ is the same as the joint distribution of the random variable $Y_{t_1+t}, Y_{t_2+t}, \cdots, Y_{t_1+t}$ for a translation of t units.

Wide smoothness is also called weak smoothness, in weak smoothness the expectation of the random variable Y_t does not change and the correlation coefficient depends on the time interval unchanged, i.e., the following conditions are satisfied:

$$\begin{cases} E\left(Y_t\right) = \mu \\ Cov\left(Y_t, Y_{t-k}\right) = r_k \end{cases} \tag{1}$$

Then the time series $\left\{Y_t, t \in T\right\}$ is wide and smooth, where the mean μ is a constant, and the autocorrelation function r_k depends only on the time lag k, independent of the time starting point.

Autoregressive model portrays the relationship between the current value and the historical value, using the historical time data of the variable's previous periods to predict the current period, i.e., to establish a linear relationship between the data of the previous periods and the data of the current period, which predicts itself, so it is called an autoregressive model. The data of the autoregressive model must satisfy the smoothness condition. The autoregressive equation depends on the historical values of the past p periods recorded as $AR(p)$ model, when the order is p, $AR(p)$ is defined as:

$$Y_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t \tag{2}$$

where φ_0 is the constant term, φ_i is the autocorrelation coefficient, ε_t is the present perturbation term, and the error term is a zero-mean white noise sequence.

$AR(p)$ has the following properties:

$$\begin{cases} E\left(\varepsilon_t\right) = 0 \\ Var\left(\varepsilon_t\right) = \sigma^2 \\ E\left(\varepsilon_t \varepsilon_s\right) = 0, \forall s \neq t \end{cases} \tag{3}$$

The moving average model is a regression of the prediction errors of the past several periods on the current period's prediction, and the current period's prediction is independent of the past period's serial values.The $MA(q)$ model establishes a linear regression of the random variable Y_t on the random perturbation term from the previous q periods.The MA model is always weakly smooth, so that the mean of the random variable Y_t is a constant, and the qth-order moving average model $MA(q)$ is expressed as follows:

$$Y_t = \mu + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{4}$$

where μ is the mean of Y_t, ε_t is the forecast error term for the current period, and ε_{t-t} is the forecast error term for the previous i periods.

$MA(q)$ satisfies the following equation:

$$\begin{cases} E\left(\varepsilon_t\right) = 0 \\ Var\left(\varepsilon_t\right) = \sigma^2 \\ E\left(\varepsilon_t \varepsilon_s\right) = 0, \forall s \neq t \\ E\left(Y_t\right) = \mu \\ Var\left(Y_t\right) = \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma^2 \end{cases} \tag{5}$$

The predicted values for the current period are not only affected by the historical values of previous periods, but also related to the previous period's disturbance term, i.e., Y_t is related to $Y_{t-1}, Y_{t-2}, \cdots$ as well as $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots$. Whether or not the data processed by the ARMA model is smooth or not depends on the data of the AR model. The (p,q) th order autoregressive moving average model $ARMA(p,q)$ is satisfied:

$$Y_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{6}$$

Where, p is the AR model order and q is the MA model order.

The ARIMA model converts the unsteady data into smooth data by one or several differencing, and processes the data in building the ARMA model. p in the $ARIMA(p,d,q)$ model is the autoregressive term, q is the MA order, and d is the number of differencing.

2.2. ARIMA modeling steps:

2.2.1. Smoothness test. The first step is to test the smoothness of the data and determine the model to be used, AR model, MA model and ARMA model are considered for smooth time series data. The unsteady time series data are transformed into smooth data after difference operation, and then transferred to the smooth data processing model. There are three methods for testing smoothness: time series plot test, autocorrelation test and unit root test. The first two methods are graphical and require knowledge of the characteristics of the smooth data in order to judge. The unit root test performs hypothesis testing by constructing a test statistic.

The mean and variance of smooth time series data are constants, and the series values fluctuate randomly up and down around a horizontal line according to a chronological plot, with no clear seasonality or trend. The smoothness of the time series data is roughly judged by observing whether the time series plot conforms to the smooth characteristics. The autocorrelation plot checks the short-term correlation of the time series data, and the autocorrelation coefficient can be judged to be smooth if it quickly recedes to zero.

If there is a unit root in the characteristic equation, the values of previous periods have a large influence on the current period, and the influence has been passed on, inconsistent with the property that the further apart a smooth time series is, the less influence it has. The ADF test is a unit root test method the original hypothesis is that there is a unit root, i.e., the series is not smooth.

2.2.2. Pure randomness test. Pure random sequence is a smooth sequence, but there is no relationship between the time series data, there is no information that can be extracted, if the sequence is a purely random sequence, the analysis stops.

2.2.3. ARMA model fixed order. In fact, ARMA model contains AR model and MA model, when the autoregressive term takes zero, ARMA model becomes MA model, and when the moving average takes zero, it becomes AR model. So directly into the ARMA model to its order. ARMA model ordering method has a variety of methods, such as according to the autocorrelation function ACF and partial autocorrelation function PACF order.

The ACF autocorrelation function is the correlation coefficient between different values of the series, reflecting the correlation of the values taken between the series. The autocorrelation coefficient for a smooth sequence delayed by k period is defined in the following equation:

$$\rho(t, t-k) = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)} \tag{7}$$

In fact there are other random variables interfering in the ACF, and the PACF erases the interferences and measures the effect between the lagged k-period series values and the current period values separately.

The Akaike Information Criterion AIC is another fixed-order method that substitutes parameters p and q to compute the AIC, prioritizing the model that achieves the minimum value. The specific calculation formula is:

$$AIC = 2n - 2\ln(L) \tag{8}$$

where L is the great likelihood function of the model and n is the number of independent parameters in the model, i.e., degrees of freedom.

2.2.4. Model testing. The test of the model includes the test of the regression equation and the test of the regression coefficient, the test of the significance of the regression equation aims at verifying the rationality of the linear equation, and the test of the significance of the regression coefficient determines whether the corresponding variable is significant to the dependent variable.

2.3. Empirical Analysis of ARIMA Predicting Opera Styles

The data in this paper is extracted from the Chinese Opera Research Center for opera styles. In this paper, the opera styles from 2008.01-2020.12 are used as the training set to construct the model, and 2021.01-2021.12 are used as the test set to verify the model. Firstly, excel is used to collect and organize the data, and r software is used to analyze and predict the data. The significance level a=0.05.

First take the year as the horizontal axis, take the monthly opera style as the vertical axis, observe the development trend of the sequence, and find that its time series in the period of 2008-2020 is a clear increasing trend, at this time we judge that the sequence is a non-stationary sequence. Since the series is not smooth, we usually use the difference or power operation to make the series smooth. Theoretically, the more times the difference is performed, the smoother the series will be after the difference. However, because each difference will lose some information, the default number of differences is not more than 2. Here we choose the first-order difference for smoothing. After the smoothing process, the sequence of opera style development does not have any increasing or decreasing trend, and it is first judged to be a smooth sequence.

To further determine whether the first-order difference series is a smooth series, in R, we introduce the ADF unit root test with the original hypothesis H0: there is a unit root. By using the adf.test() function in the tseries package in the R software, the DF value at this point is -7.1344, with a p-value of 0.01<0.05, which rejects the original hypothesis, indicating that the differenced series is a smooth series. In order to ensure that the study is meaningful, it is necessary to carry out the pure random white noise test on the differenced series. Since smooth series have short memory and the correlation usually exists only between the values of the series with a relatively short delay period, we consider the delayed LB statistic of the first 6 periods to directly determine whether the series is a white noise series. In R we utilize box.test for the test and the results show that the P-values are all much less than the significance level of 0.05, rejecting the original hypothesis, which indicates that the series is a non-white noise series and is statistically significant for the study.

According to the above description, the sequence after the first-order differencing is a smooth non-white noise sequence, next we need to determine the model. First we decompose the original data, decompose the series into trend term part, periodic part and random noise part, the original data and decomposed data are shown in Figure 1-Figure 4 respectively.

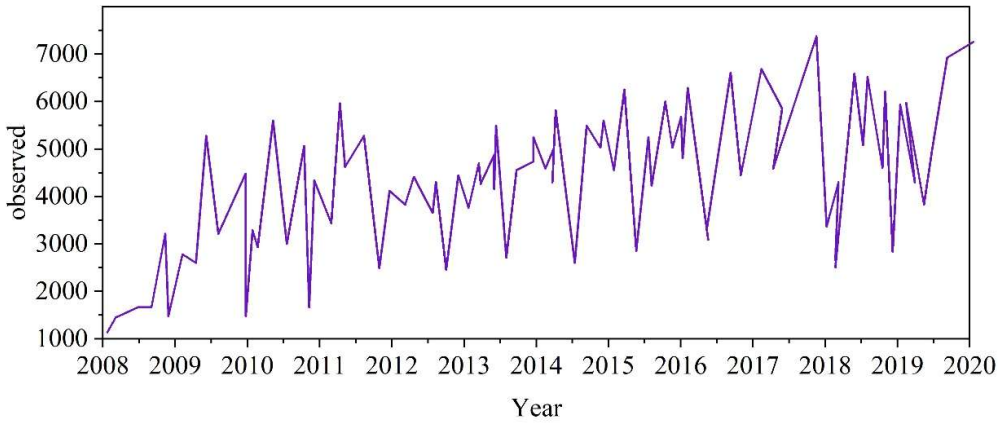

Fig. 1. Raw data

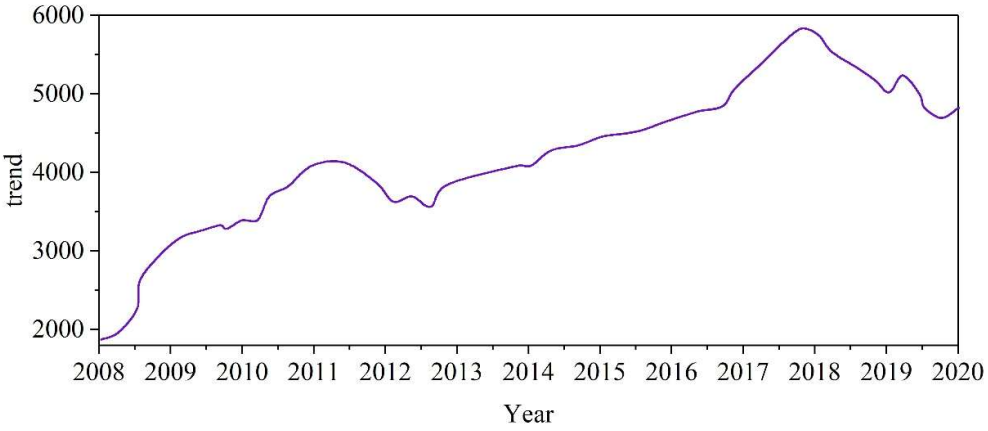

Fig. 2. Trend section

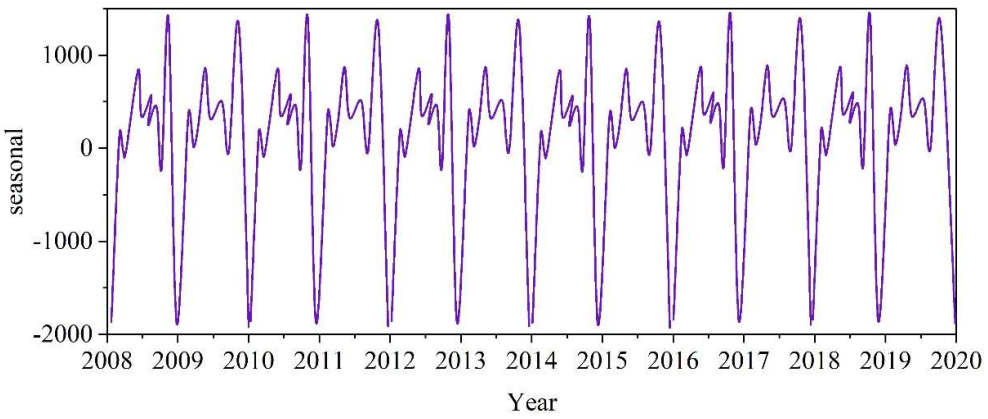

Fig. 3. Periodic part

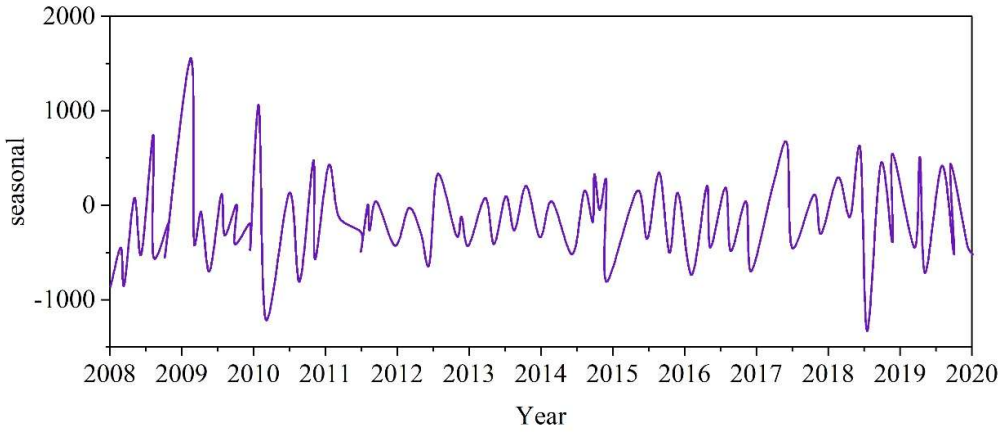

Fig. 4. Random noise section

According to the decomposition plot, we can clearly see that the series has seasonal and long-term trends, and the model is initially identified as the ARIMA seasonal product model, which is ARIMA (p, d,,q)(P, D, Q)12. Autocorrelation (ACF) and partial autocorrelation (PACF) plots are plotted for the smoothed series, as shown in Figures 5 and 6.

We can find that both ACF and PACF plots have trailing phenomenon. Since this is the graph after first-order differencing, so we can determine d = 1. Apply the least squares method to estimate the unknown parameters and select the optimal model according to the AIC criterion or BIC criterion. In R language, we use auto.arima() function to select the optimal model automatically, and the optimal model selected is ARIMA(2,1,1)(2,1,0)12 according to the principle of AIC minimization.

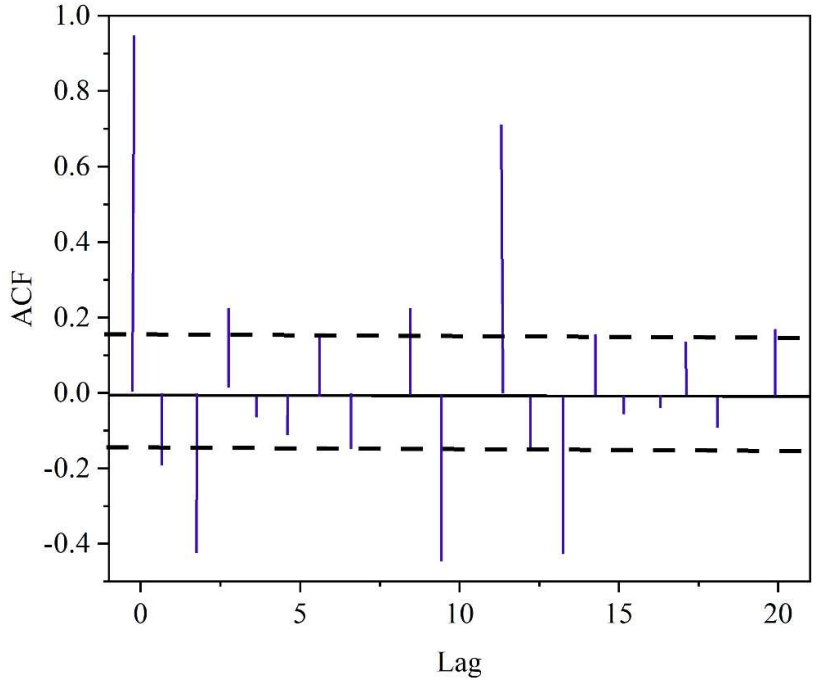

Fig. 5. Self-correlation diagram

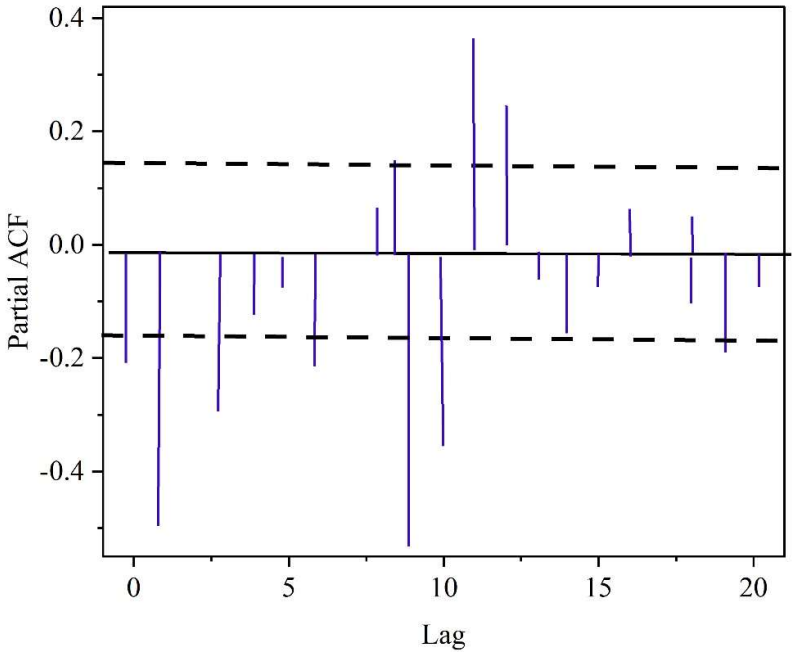

Fig. 6. Partial self-correlation diagram

Although we have obtained a specific model, we can not immediately use it, but we need to test it in many ways to ensure that the model is valid. So to test the establishment of the model, in order to ensure that the model is streamlined, we need to test the significance of the parameters of the model, the test results are shown in Table 1. The significance test of the parameters is also judged by T-test. However, because the R language does not directly provide the function of the significance test of the parameters of the time series model, so we can only through the parameter estimates/standard deviation construction of the T-statistic obtained to calculate the P-value of the test. We can see that the P-value of all parameters is much smaller than the significance level of 0.05, rejecting the original hypothesis, indicating that the parameters pass the significance test.

Table 1. Parameter significance test

Parameter	Estimate	Pr(>|z|)
ar1	0.3984	2.544518e-02
ar2	0.2546	1.294548e-01
ma1	-0.9456	1.724534e-09
sar1	-0.4851	2.703418e-05
sar2	-0.3451	8.220145e-04

Residual white noise test is a very important mitigation in the application of time series, which in order to ensure that the model has fully extracted the relevant effective information, residuals do not exist effective information. In R language, we carry out the Ljung-Box test on the model residuals, and the test results show that the P value is 0.756, which is greater than the significance level of 0.05, and the original hypothesis is accepted, which indicates that at this time, the residual series is a white noise series, which means that the model has sufficiently extracted the relevant information.

In the assumptions of the model, we believe that the random error is homoskedastic, if the residuals of the model exists heteroskedasticity situation, there is a phenomenon of aggregation of variance, it is considered that the model of the residuals of the variance of the et has an autoregressive relationship, it is necessary to construct an autoregressive relationship of the residuals of the et has an autoregressive relationship, it is necessary to construct the residuals of the ARCH or GARCH model. The residual plot is shown in Fig. 7, and the residual plot fluctuates up and down around the constant

0, and there is no residual aggregation. In order to accurately characterize the existence of heteroscedasticity of and LM test. The results show a P value of 0.3547 for the Portman_teau Q test and 0.3605 for the LM test, which are both greater than 0.05, accepting the original hypothesis, which indicates that there is no heteroskedasticity in the residuals, and that there is no need to treat the residuals any further. Therefore ARIMA(2,1,1)(2,1,0)12 was finally selected to realize the prediction of opera style.

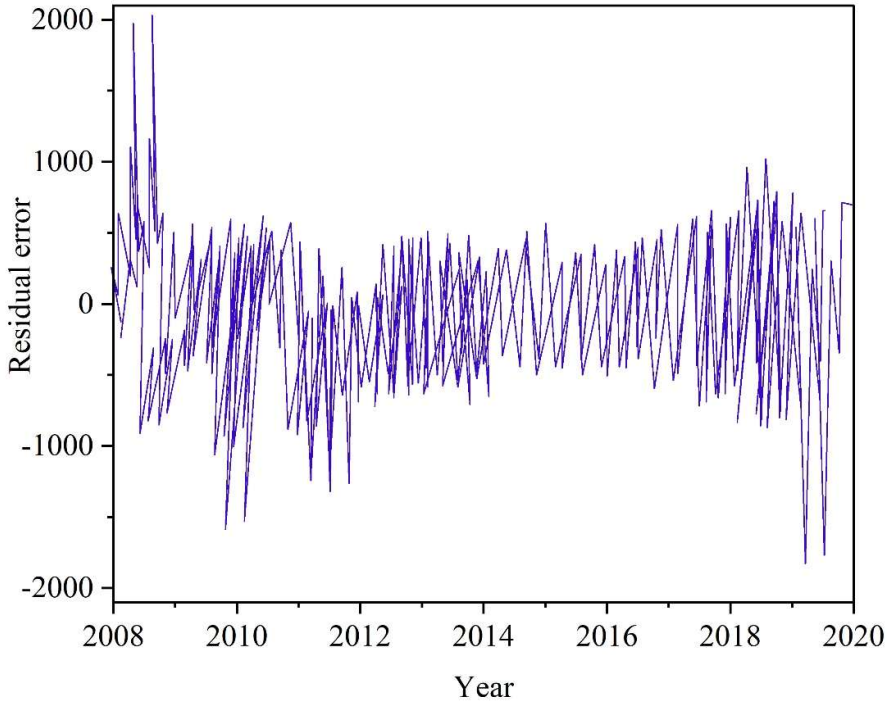

Fig. 7. Residual map

3. Automatic music generation method based on generative adversarial network

In recent years, deep neural networks such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) have been applied to the field of music generation and have shown great potential for their generative capabilities. Based on the understanding of how popular music is composed, three different approaches combine GANs to process the interactions between tracks and generate polyphonic music with harmonic and rhythmic structures. Adversarial training of continuous recurrent networks leads to highly flexible and expressive models with the ability to control pitch height, frequency, intensity, and duration of fully continuous sequential data, but still falls short for music generation of longer sequences.

However, GAN methods are still limited, and the research on music generation is still in its infancy, and their research fails to fully reflect the complex relationship between the musical elements in the spatial and temporal dimensions of real musical compositions. Moreover, compared to text generation, music generation deals with longer time sequences, and the network is prone to fail to notice the correlation before and after the long sequences when learning, so to address the above problems, this chapter investigates a generative method for music text vignettization.

Inspired by the GAN approach, this chapter chooses the Leak-GAN model to generate music.The Leak-GAN model consists of a discriminator and a generator, which allows the discriminator to leak its extracted features from the upper level to the generator to help the learning of more distant sequences, thus solving the problem of long sequences in music generation to a certain extent. Next, we first

introduce the general concept of generative adversarial networks, and then describe the Leak-GAN model in detail.

3.1. Generating Adversarial Networks

Generative Adversarial Networks (GAN) [17] is an unsupervised learning method designed to learn by having two neural networks compete with each other. While one of the neural networks, called Generator (G), maps random noise z sampled from a prior distribution to the data space, the other neural work, called Discriminator (D), outputs a single scalar that represents the probability that the training examples come from the data, rather than the generator's distribution. The two networks train each other through round after round of generation and discrimination to outperform each other simultaneously. The adversarial learning process can be described as a two-player minimax game between a generator and a discriminator with an objective function:

$$\begin{aligned}\min_G \max_D V(D,G) = & E_{x\sim p_{data}(x)}\left[\log D(x)\right] \\ & + E_{z\sim p_z}\left[1-\log(D(G(z))\right]\end{aligned} \tag{9}$$

where P_{data} is the distribution of the true data x and P denotes the prior distribution of z.

3.2. Leak-GAN

Leak-GAN, is mainly to leak the high-level discriminator feature information to the manager module, which uses LSTM [18] network to act as an intermediate medium to form the behavioral target output with guidance, while the underlying generator inputs word vectors, which are passed through the worker module to get the behavioral vectors, and combines with the target output sampling to get the output vectors of the next word. The so-called Leak, which leaks the feature information of the discriminator into the generator and guides the generator generation, optimizes the traditional adversarial model.

For automatic music generation, the text representation melody approach studied in this paper allows music generation to be viewed as a one-dimensional sequence generation process. In each time step t, the current state of the model is denoted as $s_t = (x_1, \ldots, x_i, \ldots, x_t)$ and x_{t+1}, respectively, where x_i is a candidate token in a given vocabulary V. Improvement of the generator G_θ is guided by training the θ parameters in the generator G_θ and the ϕ parameters in the discriminator D_ϕ, and using the discriminator D_ϕ.

After allowing leakage of the high-level feature representation of the discriminator D_ϕ to the generator G_θ, which should not receive such information, the Leak-GAN model solves the problem that the bootstrap signals of D_ϕ are not fully taken into account by the signals when generating the entire melodic sequence S_T. Specifically, as the length of the generated sentence T becomes longer, the bootstrap signal of a single scalar that acts as a discriminator usually contains less information. To alleviate this situation, the Leak-GAN model not only provides discriminator information to the generator, but also applies a hierarchical reinforcement learning architecture to the generator in order to give G_θ the leaked information.

The manager and worker modules of the generator are first introduced, both of which start from an all-zero hidden state, denoted as h_0^W and h_0^H. Figure 8 shows an overview of the model structure, where in each step the discriminator sends the leaked feature vector f_t to the manager and generates the target vector g_t in conjunction with its current state:

$$\hat{g}_t, h_t^M = M\left(f_t, h_{t-1}^M; \theta_m\right) \tag{10}$$

$$g_t = \frac{\hat{g}_t}{\left\|\hat{g}_t\right\|} \tag{11}$$

where $M(\cdot;\theta_m)$ denotes the LSTM module used by the manager, the current hidden vector is denoted as h_t^W, and the leaked feature vector is f_t.

$$D_\phi(s) = sigmoid\left(\phi_f F\left(s,\phi_f\right)\right) = sigmoid\left(\phi_f f\right) \tag{12}$$

where $sigmoid(z) = 1/\left(1+e^{-z}\right)$, $F\left(\cdot;\phi_f\right)$ denote the feature extractors of the CNN and $f = F\left(s;\phi_f\right)$ is the feature vector of the last layer in D_ϕ.

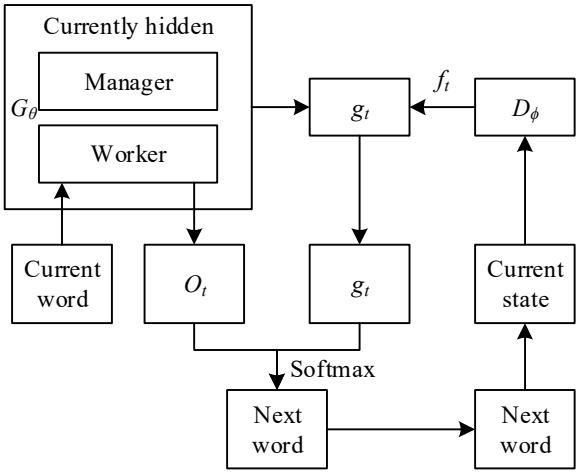

Fig. 8. Leak-GAN model framework

Next is the current sum of c targets generated with the weight matrix W_ψ, where ψ is linearly varying and the target embedding vector w_t is defined as follows:

$$w_t = \psi\left(\sum_{i=1}^{c} g_{t-i}\right) = W_\psi\left(\sum_{i=1}^{c} g_{t-i}\right) \tag{13}$$

The current note x_t of the generated melody is used as input to the worker module. In order to obtain the final spatial distribution of the next note to be selected, a matrix product is realized from O_t over w_t by softmax in the current state of the manager:

$$O_t, h_t^W = W\left(x_t, h_{t-1}^W;\theta_w\right) \tag{14}$$

$$G_\theta\left(\cdot \mid s_t\right) = soft\max\left(O_t \cdot w_t / \alpha\right) \tag{15}$$

where h_t^W is a cyclic hidden vector of the LSTM, $W(\cdot;\theta_w)$ denotes the worker module. O_t denotes the vector matrix of all notes, and α is a temperature parameter that indicates the generated entropy.

4. Experimental results of musical melody generation

Regarding the evaluation indicators, this experiment was planned to use BLEU at the beginning. However, it was found in the experiments that the BLEU values were not effective in assessing the quality of the generated melodies. Melodies that sounded poorly paced obtained high BLEU values. Moreover, the BLEU values obtained by all models were high and very close to each other, without differentiation. Therefore, this experiment decided to evaluate the models by comparing the statistics on the generated sample set with those on the training set. The more similar to the statistics on the training set, the better the model. The statistics used here are, SPB and QN, which represent the musical span of each measure and the percentage of "qualified notes" to the total number of notes, respectively ("qualified notes" are those notes that exceed the set duration). SPB is able to characterize

the degree of melodic variation, and QN is able to effectively detect whether the generated melody is too fragmented or not, and they are calculated as follows:

$$SPB = \frac{\sum Octave\ span\ per\ bar}{Total\ number\ of\ bars\ in\ the\ melody} \tag{16}$$

$$QN = \frac{Number\ of\ notes\ that\ match\ the\ set\ duration}{Notes\ of\ the\ melody} * 100\% \tag{17}$$

4.1. Music Melody Generation Effect

Normal GAN, MusicGAN, Leak-GAN_1 and Leak-GAN_2 are compared in the experiment, and Table 2 shows their effects. Among them, for QN this experiment is set up with three durations, they are 4x, 6x and 8x sixteenth notes. The second row of the table gives the values of Nottingham dataset on SPB, QN statistics. Comparing this row with other rows, we can see that Leak-GAN (including Leak-GAN_1 and Leak-GAN_2) is much better than normal GAN, which shows that Leak-GAN can indeed effectively reduce the bias of GAN optimization goals and thus improve the generation quality. Leak-GAN can be regarded as a combination of GAN and LSTM, which increases the adversarial process and thus the generative power compared to ordinary GAN. From this perspective, it also explains why Leak-GAN can greatly outperform ordinary GAN. In addition, Leak-GAN outperforms MusicGAN, and the PB value of Leak-GAN_2 is 0.12 higher than that of MusicGAN, which suggests to some extent that Leak-GAN alleviates the "post-collapse" problem in training.

Table 2. The results of each sequence generation model are compared

Model	SPB	QN		
		4	6	8
Notingham	0.39	68.25	46.59	45.36
Ordinary GAN	0.32	76.12	54.32	54.36
Music GAN	0.29	75.14	51.24	50.23
Leak-GAN_1	0.41	78.54	57.51	56.15
Leak-GAN_2	0.41	79.12	58.12	57.64

Figure 9 gives a comparison of the interpolation generation for each sequence generation model. It can be seen that Leak-GAN_2 is comparable to MusicGAN, they are slightly better than the normal GAN, but overall they have the same variation trend. Leak-GAN_1 on the other hand has a different trend and compared to the other its variation is smaller. This may indicate that for the Leak-GAN_1 dataset the latent encoding is more densely distributed over the latent space, thus making it the best for generation.

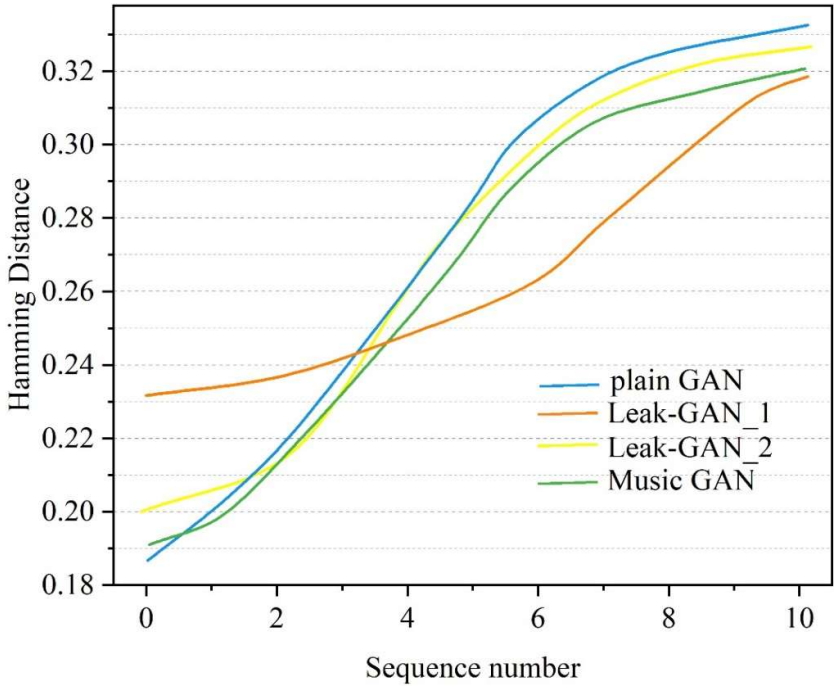

Fig. 9. The interpolation of each sequence generation model is compared

4.2. Automatic recognition and 2D visualization of loudness in musical melodies

To improve the readability of the graphic, graphic gridlines are added. Considering again that the automatically generated longitudinal gridlines cannot express the characteristics of the music, the longitudinal gridlines are plotted on a bar-by-bar basis. The graphical effect of plotting after refining the pitch multilevel scale and bar gridlines is shown in Figure 10.

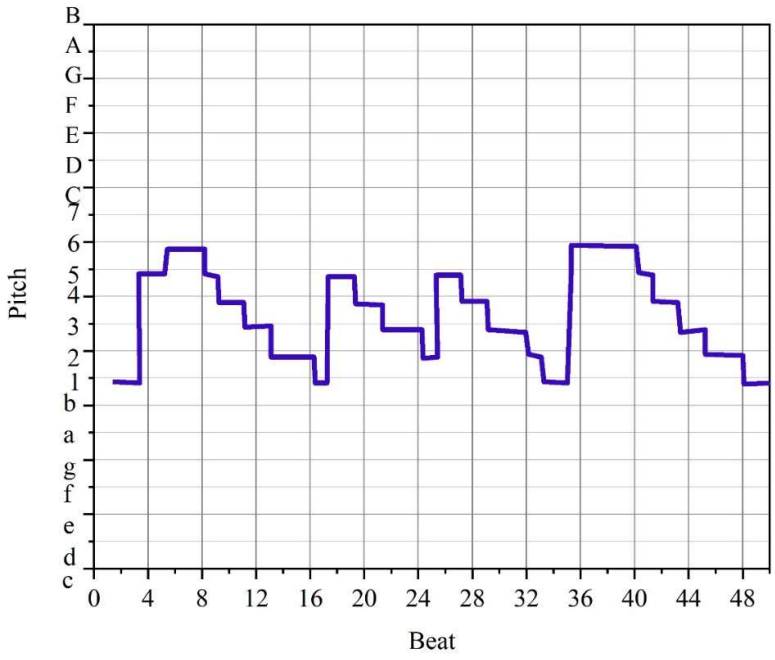

Fig. 10. The melody visualization of multistage pitch and section lines

In addition to the method of generating melodic feature matrices such as pitch and time value manually, it should also be possible to accomplish the automatic extraction of the corresponding feature values for the input music (fragment) file. Here the music file was obtained by recording a

female voice chanting. In order to supplement the missing loudness features of the melody in the previous section, so only the extraction of loudness features is involved.

The frequency composition, frequency width and other features can be analyzed using Fourier analysis. The Fourier transform transforms the traditional method of analyzing signals in the time domain into a method of analyzing problems in the frequency domain, based on the principle that any continuously measured time series or signals can be represented as an infinite superposition of sinusoidal signals of different frequencies. Because computers can only handle discrete, finite-length sequence data, the input values used by computers to perform the DFT are N specific sample values, that is, the signal values in the time domain, and the number of input sample points, N, determines the computational scale of the transformation. The Fast Fourier Transform is a fast algorithm for the Discrete Fourier Transform, which can speed up the processing speed of the computer. According to the sampling theorem, the highest frequency that can be resolved by FFT is half of the sampling frequency, and the function FFT return value is symmetric on the axis of Nyquist frequency. The spectrogram of the signal after FFT can be used N/2 + 1 points of information, the X-axis corresponds to the frequency sequence of the time series, the maximum frequency of the X-axis is FS/2, the Y-axis is the amplitude, is the mode of the complex number obtained after FFT. In order to distinguish the different frequency components, the sampled data points N should take large enough values. For music (clips), the time length is a constant value TS and the sampling frequency is 44100Hz, so the total number of sampling points N is T*44100. The music samples used here. The plotted spectrum is shown in Figure 11. Zooming in on the original graph, it can be seen that the frequency values with higher amplitudes are in the range of 200-400 Hz, and the enlarged image is shown in Figure 12. The frequency values corresponding to the highest amplitude values were further obtained, i.e., the pitch at which the loudness enhancement was determined. For the music samples used here, the amplitude values of 8496th and 1187544th reached a maximum of 18459 Hz. Considering the symmetry of the FFT, the frequency value of 8496th ordinate is taken and the corresponding frequency value is 315.69 Hz.

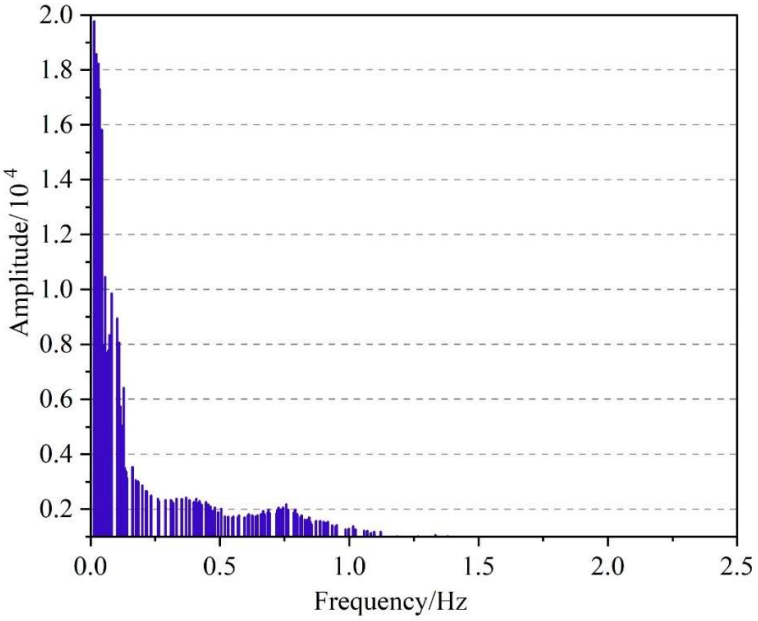

Fig. 11. The spectrum of the Nyquist frequency

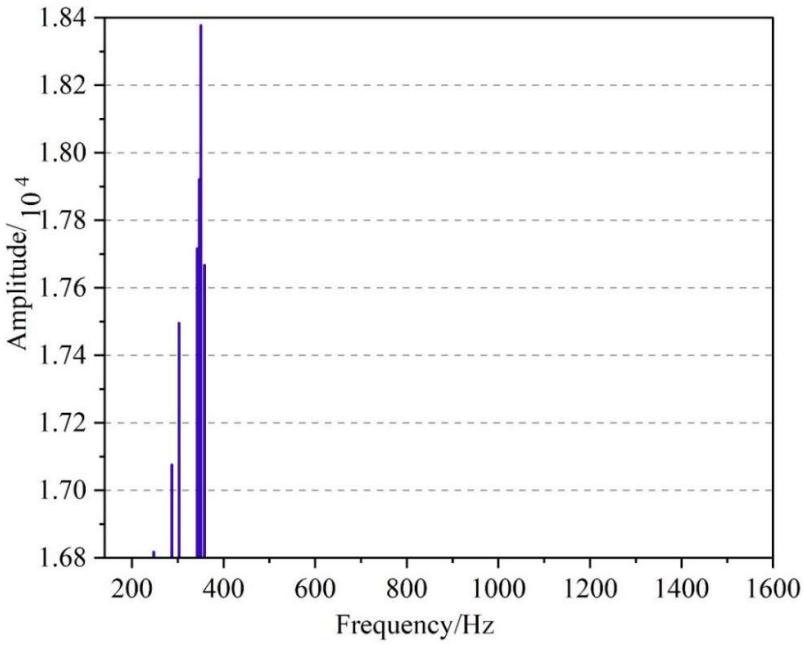

Fig. 12. A larger part of the spectrum

The size of the loudness is closely related to the individual interpretation design, and since the fundamental frequency of the human voice varies greatly, the fundamental frequency of a single tone should be determined first. Using the same method described above to extract frequency features for the do tone alone, here the frequency of the music sample do tone is 186.45Hz, and it can be determined that the frequency value with the largest amplitude of 315.69Hz corresponds to the la tone. The line pattern of the la tone is plotted as a continuous time-value length line with "red triangles", so as to increase the plotting elements that express the loudness characteristics, and Fig. 13 shows the melodic ladder diagram that includes the loudness characteristics.

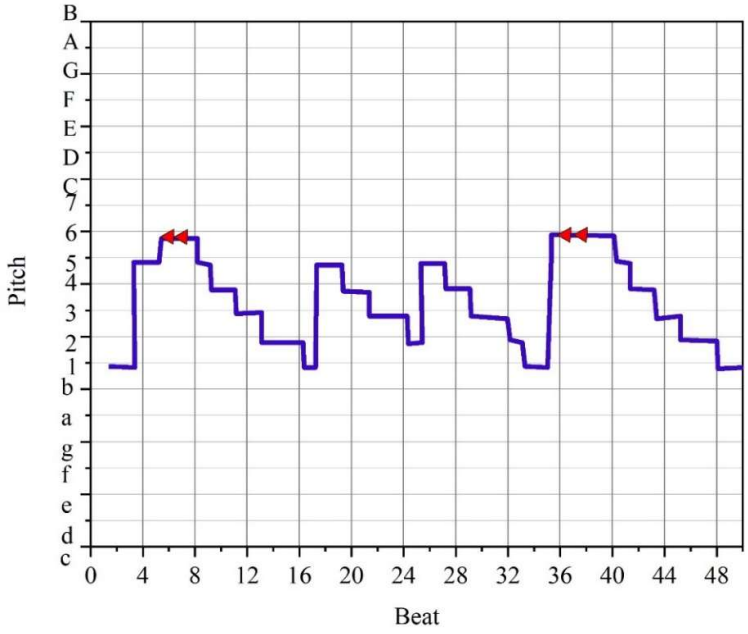

Fig. 13. The melody that contains the characteristics of the sound

5. Conclusion

In this paper, an opera style prediction model based on ARIMA model and a multimodal generation strategy of music melody based on generative adversarial network are established, and a training set is built to train the models.

The time series of opera style development shows that the Chinese opera style is in the trend of year-by-year evolution, and this paper chooses the first-order difference to smooth the time series of opera style development, and the processed sequence is a non-white noise sequence, which is statistically significant for research. By determining the model order, parameter estimation, multiple tests of the model and other steps, the model suitable for predicting the time series of the development of opera styles is determined as ARIMA(2,1,1)(2,1,0)12 model.

The Leak-GAN model increases the adversarial process compared to the ordinary GAN, which improves the model generation ability. In addition to this, the Leak-GAN model outperforms MusicGAN, and the PB value of Leak-GAN_2 is improved by 41.38% compared to MusicGAN. It shows that the Leak-GAN model alleviates the “post-collapse” problem in model training, and the Leak-GAN model has the best effect on the multimodal generation of music melody.

By extracting the features of the music melody through the Fourier transform, and mapping the features of the music melody into the two-dimensional visualization effect graph, we can not only carry out the two-dimensional visualization analysis of the frequency composition, frequency width and other features, but also visualize the music melody containing loudness features, and determine that the la tone corresponds to the frequency value of the largest amplitude of 315.69 Hz. The method of this paper has an irreplaceable role in improving people's music appreciation and enhancing the quality of opera style and music melody generation.

Funding

1)Major Scientific Research Projects of Higher Education Institutions in Anhui Province: Research on the Era Characteristics and Inheritance of the Dabie Mountain Spirit in Wanxi Lu Opera (2023AH040341);
2) Anhui Province Social Science Innovation and Development Research Project: Research on the Historical Changes and Inheritance of the Era Cognition and Presentation of Lu Opera on the West Road (2022CX149);
3)Anhui Provincial Education Science Research Project Special Project: Research on Quality Monitoring of Music Curriculum Implementation in Compulsory Education Stage in Lu'an Region (JKT24145);
4)Key Scientific Research Project of Universities in Anhui Province: Research on the Aesthetic Thoughts in the Qing Dynasty Folk Music "Liaozhai Folk Qu" (2024AH053470).

References

[1] Wu, J., Jiang, K., & Yuan, C. (2019). Determinants of demand for traditional Chinese opera. Empirical Economics, 57, 2129-2148.

[2] Lin, Y., & Colbert, F. (2018). Chinese opera and the international market. International Journal of Arts Management, 75-82.

[3] Li, T., & Gong, L. (2024). Cultural Identity and Historical Narration: An Analysis of the Expression of National Consciousness in Chinese Ethnic Opera Works. Revista Electronica de LEEME, (54).

[4] Rao, N. Y. (2017). Chinese Opera Percussion from Model Opera to Tan Dun. China and the West: Music, representation, and reception, 163-185.

[5] Yating, W. U., RAHMAN, A. R. A., KMW, P., & LING, S. M. (2021). The origin and formation of sichuan operA. Journal of the Balkan Tribological Association, 27(2).

[6] Leung, B. W. (2023). Transformation of traditional art forms in the evolving contexts: Cantonese opera in Hong Kong as an example. Cultural Sustainability and Arts Education: International Perspectives on the Aesthetics of Transformation, 53-63.

[7] Shuai, L., & Yodwised, C. (2024). " Hometown is Beijing" The Elements of Learning Peking Opera with the Style of Beijing Culture in Modern Songs. Journal of Modern Learning Development, 9(5), 356-361.

[8] Xiao, X. (2023). The Use of Traditional Opera Elements in the Characterization Process of Contemporary Stage Drama. Highlights in Art and Design, 4(1), 13-16.

[9] Han, Q., & Zhang, R. (2017). Acoustic analyses of the singing vibrato in traditional Peking opera. Journal of Voice, 31(4), 511-e1.

[10] Lam, J. S. (2022). Kunqu: A classical opera of twenty-first-century China. Hong Kong University Press.

[11] Jiang, R., & Jamalludin, N. I. (2023). Modern Chinese Opera in the formative years: Extrapolation of music and its socio-cultural development. Journal in Advanced Humanities, 4, 4.

[12] Yue, X. (2024). Comparative Analysis of the Differences in Aria Melodies at Different Stages of Chinese National Operas. Journal of Education, Humanities and Social Sciences, 26, 401-408.

[13] Chen, J. (2024). Aesthetic Value and Cultural Philosophy in Opera: A Cross-Era Perspective. Cultura: International Journal of Philosophy of Culture and Axiology, 21(3).

[14] Zhao, Y., Cheng, B., Huang, Y., & Wan, Z. (2023). Beyond Words: An Intelligent Human‐Machine Dialogue System with Multimodal Generation and Emotional Comprehension. International Journal of Intelligent Systems, 2023(1), 9267487.

[15] Liu, P. (2022, January). The decline of traditional chinese opera. In 2021 International Conference on Public Art and Human Development (ICPAHD 2021) (pp. 878-881). Atlantis Press.

[16] Vipasha Sharma,Swagata Ghosh,Varun Narayan Mishra & Pradeep Kumar. (2025). Spatio-temporal Variations and Forecast of PM2.5 concentration around selected Satellite Cities of Delhi, India using ARIMA model. Physics and Chemistry of the Earth103849-103849.

[17] Guodong Jin,Yuxiang Liu,Ran Wei,Bining Yang,Bo Pang,Xinyuan Chen... & Kuo Men. (2025). Prediction of real-time cine-MR images during MRI-guided radiotherapy of liver cancer using a GAN-ConvLSTM network.. Medical physics

[18] Ziwei Pan,Lei Xu & Nengcheng Chen. (2025). Combining graph neural network and convolutional LSTM network for multistep soil moisture spatiotemporal prediction. Journal of Hydrology132572-132572.