J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 4187--4211

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

Combinatorial Press

A Study of Piano Tone Teaching in the Context of Artificial Intelligence Interaction

Ruijie Liao[1,✉]

[1] *Xiamen University Tan Kah Kee College, Zhangzhou, Fujian, 363105, China*

ABSTRACT

Piano timbre recognition and intelligent synthesis are of great significance in realizing the intelligent teaching of piano timbre. This paper takes the piano timbre teaching based on artificial intelligence interaction as the research object, constructs the timbre expression spectrum based on harmonic structure through the exploration of timbre synthesis, timbre features and other related theories, proposes the timbre feature extraction method based on the time-frequency cepstrum domain of the piano music signal, and then constructs the piano timbre recognition and intelligent synthesis system, realizes the simulation of the piano music, and then provides an intelligent interactive tool for the piano timbre teaching. The method is used to construct a piano tone recognition and intelligent synthesis system. When using the method in this paper, the amplitude of the piano tends to be stable when the frequency is 1600Hz~2400Hz, and there is no noise interference, and when the frequency is 2500Hz and 2800Hz, the amplitude is the lowest, and the recognition performance of the piano timbre is better. Meanwhile, the correct rate of timbre recognition of this method reaches 87.83%, which is better than 58.54% of the comparison method. In addition, the musical tone signals simulated by the method in this paper are very close to the theoretical values of each note of the real piano instrument captured, with an accuracy rate of up to 99%, which proves the accuracy of the simulated piano sounding. And the method can effectively promote the combination of artificial intelligence technology and piano teaching concept, the confidence level of quantitative regression analysis is high, and the evaluation results of teaching quality are good, which provides a reliable theoretical and practical basis for realizing the high-quality teaching of piano timbre.

Keywords: Musical sound simulation, harmonic structure, time-frequency cepstrum domain, piano timbre recognition, artificial intelligence

1. Introduction

The word timbre has many different interpretations and expressions. "The Concise Oxford Dictionary

✉ Corresponding author.
E-mail address: italy3493@163.com (R. Liao).
Received 05 March 2024; Revised 20 May 2024; Accepted 05 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-231

of Music states that timbre refers to the characteristics of sounds produced by different instruments or vocals. Examples include the difference between a flute and a clarinet, and the difference between a soprano and a mezzo-soprano [1-2].", "The American National Standards Institute suggests that timbre is a certain property of a sound produced auditorily by which the listener is able to judge the difference between two sounds presented in the same way and having the same pitch and loudness [3-4]." Timbre as an acoustic phenomenon is able to accurately demonstrate the different tonal qualities and colors of sounds. When used in music composition and performance, it is capable of providing a unique aesthetic experience [5-6].

Because of the piano's beautiful sound, excellent performance, all-purpose features known as the "king of musical instruments", the modern piano 88 keys cover almost all the ranges used in music, can be a single melody or polyphonic performance, both solo and a strong accompaniment, but also educators in music teaching the most commonly used, most visual and intuitive teaching aids [7-9]. Therefore, more and more people from all walks of life love and learn the piano in modern society, and they take learning to play the piano as a way to improve their quality and cultivation [10-11].

In modern music colleges and universities, piano is also listed as a specialized subject major, which gives people who love it a chance to learn it in a scientific, formal, systematic and comprehensive way. As a performing art, the primary element of piano performance is timbre, melody is the foundation and primary condition of music creation, while timbre is the first element in music performance [12-14]. Imagine an excellent musical work without pure, beautiful tone as a carrier, it is doomed to failure, it can not provoke people's appreciation of interest and resonance. So the first element of piano performance is good tone [15-16].

In the teaching of piano performance art, tone training can be said to be an important link throughout. The interpretation of the timbre of the piano works mainly comes from the imagination and control of the performer on the sound color of the works. That is to say, when interpreting the timbre of a piano work, the performer needs to make comprehensive use of his or her two abilities - timbre imagination and timbre control [17-19]. Obviously, among these two abilities, the former belongs to the thinking level, while the latter belongs to the technical level. Then, it is worth thinking about how teachers should develop students' thinking ability while cultivating their playing (technical) ability in actual teaching, so as to improve their timbre interpretation level of piano playing [20-21].

This paper firstly elaborates the basic theory of timbre synthesis and the relevant features of timbre, and then proposes a timbre feature extraction method for piano music signal on this basis. Specifically, by preprocessing the piano music signal, we constructed the timbre expression spectrum based on harmonic structure, and constructed the timbre feature extraction system based on the time-frequency cepstrum domain of the piano music signal, so as to realize the intelligent identification of the piano timbre and the intelligent simulation of the music, and thus constructed the piano timbre teaching mode based on the interaction of artificial intelligence. Finally, in order to verify the effectiveness of the method, this paper carries out system testing, piano music simulation experiments and the evaluation of the effect of its use in piano timbre teaching.

2. Piano Tone Teaching Based on Artificial Intelligence Interaction

This chapter firstly introduces the theories such as the basis of timbre synthesis and timbre features, and then proposes a timbre feature extraction method for piano music signals as a way to realize piano timbre teaching based on artificial intelligence interaction.

2.1. Fundamentals of Tone Synthesis

2.1.1. Equation of vibration of a string. Piano articulation principle: the basic sound source of the piano is the strings, the player presses the keys to trigger the hammer to strike the strings, so that the strings vibrate to produce the original musical sound, followed by the resonance system to produce resonance, the radiation system will be a large area of sound waves radiated out, so that the sound color to be touched up, the volume of sound to be enhanced. The vibration of piano strings is analyzed as follows.

Assuming that in piano playing, the hammer strikes the string at a fixed application position, with a short contact time and a fixed and small hammer width, let the length of the string be l, the ends fixed, the width of the hammer 2δ, and the position of the point of strike $x = x_0$, it can be seen that the range of strikes X is $x_0 - \delta \le X \le x_0 + \delta$.

Let the initial velocity of the piano vibration is v_0, the acceleration is a, according to the time of t. Assuming that the mallet strikes the string at the time of striking, the striking point $x = x_0$ obtains the maximum initial velocity v_0, the striking range boundary point $x_0 - \delta$ and $x_0 + \delta$ obtains the initial velocity of 0, and the strings are free to vibrate after the end of the striking, then the equation of the vibration of the strings is expressed as follows:

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} = a^2 \dfrac{\partial^2 u}{\partial x^2} \\ u\big|_{x=0} = u\big|_{x=l} = 0 \\ u\big|_{t=0} = \varphi(x) = 0 \\ \dfrac{\partial u}{\partial t}\Big|_{t=0} = \begin{cases} v_0 \cos \dfrac{x - x_0}{2\delta}\pi, (x_0 - \delta \le x \le x_0 + \delta) \\ 0, 0 \le x \le x_0 - \delta \\ 0, x_0 + \delta \le x \le l \end{cases} \end{cases} (0 < x < l) \tag{1}$$

From the boundary conditions, the eigenfunction is $\sin \dfrac{n\pi x}{l}$. The general equation for the vibration of a piano string is:

$$u(x,t) = \sum_{n=1}^{\infty} \left(C_n \cos \frac{n\pi a}{l} t + D_n \sin \frac{n\pi a}{l} t \right) \sin \frac{n\pi}{l} x \tag{2}$$

The string vibration equation (2) is substituted into the initial conditions to obtain the coefficients C_n and D_n:

$$\varphi(x) = \sum_{n=1}^{\infty} C_n \sin \frac{n\pi}{l} x = 0 (0 < x < l) \tag{3}$$

$$\psi(x) = \sum_{n=1}^{\infty} D_n \frac{n\pi a}{l} \sin \frac{n\pi}{l} x = \begin{cases} v_0 \cos \dfrac{x - x_0}{2\delta}, (x_0 - \delta < x < x_0 + \delta) \\ 0, (0 < x < x_0 - \delta) \\ 0, (x_0 + \delta < x < l) \end{cases} \tag{4}$$

In Eqs. (3) and (4), the left side of the equal sign is the Fourier series of the function on the right, which can be derived:

$$\begin{cases} C_n = 0 \\ D_n = \dfrac{2}{n\pi a}\displaystyle\int_{x_0-\delta}^{x_0+\delta} v_0 \cos\frac{\xi - x_0}{2\delta}\pi \sin\frac{n\pi\zeta}{l}\, d\zeta \\ = \dfrac{2v_0}{n\pi a}\displaystyle\int_{x_0-\delta}^{x_0+\delta}\left[\cos\frac{\pi x_0}{2\delta}\cos\frac{\pi\zeta}{2\delta} + \sin\frac{\pi x_0}{2\delta}\sin\frac{\pi\zeta}{2\delta}\right]\sin\frac{n\pi\zeta}{l}\, d\zeta \\ = \dfrac{4v_0\delta}{n\pi^2 a}\left[\dfrac{1}{1-\frac{2\delta n}{l}}\sin\frac{n\pi x_0}{l}\cos\frac{n\pi\zeta}{l} + \dfrac{1}{1+\frac{2\delta n}{l}}\sin\frac{n\pi x_0}{l}\cos\frac{n\pi\zeta}{l}\right] \\ = \dfrac{4v_0\delta}{n\pi^2 a}\left[\left(\dfrac{1}{1-\frac{2\delta n}{l}} + \dfrac{1}{1+\frac{2\delta n}{l}}\right)\sin\frac{n\pi x_0}{l}\cos\frac{n\pi\zeta}{l}\right] \\ = \dfrac{8v_0\delta}{n\pi^2 a}\dfrac{1}{1-\frac{4\delta^2 n^2}{l^2}}\sin\frac{n\pi x_0}{l}\cos\frac{n\pi\zeta}{l} \end{cases} \tag{5}$$

Thus, a representation of the string vibration equation for a piano can be derived as follows:

$$u(x,t) = \frac{8v_0\delta}{\pi^2 a}\sum_{n=1}^{\infty}\left(\frac{1}{n}\frac{1}{1-\frac{4\delta^2 n^2}{l^2}}\sin\frac{n\pi x_0}{l}\cos\frac{n\pi\delta}{l}\sin\frac{n\pi a t}{l}\sin\frac{n\pi x}{l}\right) \tag{6}$$

From equation (6) of the string vibration of a piano:

1) The waveform of a piano string vibration is a simple summed superposition of the fundamental and all the overtones, the amplitude of which varies from one overtone to another.

2) Among all waveforms, the fundamental (corresponding to $n = 1$) has the greatest energy, so the human auditory system is most sensitive to the fundamental.

3) In the waveform of piano string vibration, the energy of the fundamental and all overtones decays proportionally to $\frac{1}{n}$, while the energy of all overtones of plucked instruments (e.g., guitars, etc.) decays proportionally to $\frac{1}{n}$. Therefore, for the human auditory system, the sound of a piano contains more overtones that can be perceived by human beings (in the range of the human auditory frequency), and thus the tone color is richer and fuller.

2.1.2. Tone composition analysis. Piano strings vibrate in the form of transverse, longitudinal, octave, and torsional vibrations. Full-length transverse vibration and segmental transverse vibration produce the fundamental and a series of overtones. The transverse vibration is the largest in amplitude, determines the pitch, and forms the basis of the timbre. Periodic changes in string length and tension cause longitudinal vibration, which has a much higher frequency than the fundamental frequency and also produces a series of harmonics. Doubling frequency vibration is generated by the non-absolute fixation of the hanging point of the string, the frequency is twice the fundamental frequency of the transverse vibration, and there is a series of harmonics. String torsional vibration is produced by plucking and rubbing the strings, the fundamental frequency is lower than the transverse vibration, and there are series of harmonics. It can be seen that piano string vibration produces a rich harmonic series.

The timbre of piano audio does not only depend on the fundamental and overtones, but all frequency components in the frequency domain have some influence on the timbre formation. Theoretically, since the musical sound is compounded by overtones of different frequencies and different intensities (here, all frequency components including harmonic series), any change in frequency components and their intensities will cause changes in the frequency and time domains, i.e., changes in their spectral characteristics as well as their envelope characteristics, which will ultimately affect the change of timbre. Therefore, the reconstruction of the piano audio cannot only use the frequency components of the fundamental and overtones, but use all the frequency components in order to reproduce the original timbre.

2.2. Tonal characteristics

The transformation from digitized musical time waveform signals to concise and efficient timbre features is the key to instrument timbre recognition. In this paper, piano timbre features are roughly divided into three categories: time domain features, frequency domain features and cepstrum features.

2.2.1. Time domain characteristics. Time-domain features are the most intuitive and describe the temporal waveform information of a musical tone. As a member of timbre features, its important role cannot be ignored.

1) Zero Crossing Rate. Zero Crossing Rate (ZCR) calculates the number of times a signal crosses the time axis in each frame. To a certain extent, it reflects the nature of the spectrum and can be regarded as a rough estimate of the frequency. The overall performance is small when there are many periodic components and large when there are many noise components. The over-zero rate can be calculated according to equation (7):

$$N_{ZCR} = \sum_{n=0}^{N-2} \left| \operatorname{sgn}[x(n)] - \operatorname{sgn}[x(n+1)] \right| \tag{7}$$

where N denotes the number of points in a frame of signal, $x(n)$ denotes a frame of signal, and $\operatorname{sgn}[\cdot]$ denotes the sign function, i.e:

$$\operatorname{sgn}[x(n)] = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases} \tag{8}$$

There are some differences in over-zeroing rates between instruments, especially between harmonic and non-harmonic instruments. Although the over-zero rate alone is not superior in timbre differentiation, the over-zero rate can be used as a complementary feature to the feature set to achieve better results.

2) Autocorrelation coefficient. The autocorrelation coefficient (AC) is a time-domain representation of the spectral distribution, which is the inverse Fourier transform of the signal power spectrum. It is expressed by equation (9):

$$A_{AC}(k) = \frac{1}{A_{AC}(0)} \sum_{n=0}^{N-k-1} x(n) \cdot x(n+k) \tag{9}$$

For timbre differentiation, the autocorrelation coefficient proved to be a very valuable feature.

3) Logarithmic Start Time. The logarithmic onset time (LAT) refers to the duration of the onset segment and can be calculated using equation (10):

$$T_{LAT} = \log_{10}\left(t_{end} - t_{st}\right) \tag{10}$$

In this case, t_{end} refers to the ending moment of the initial section, and t_{st} refers to the beginning moment of the initial section.

The onset of the instrument plays a key role in the perception of the timbre. Although the logarithmic start time is only a simple description of the length of the start, it still performs well.

2.2.2. Frequency domain characteristics. Frequency domain features mostly characterize the amplitude spectrum in terms of a probability distribution. The calculation of frequency domain features is based on the assumption that the normalized amplitude spectrum can be regarded as some kind of probability distribution of sound energy with frequency. Frequency domain features include spectral center of mass, spectral standard deviation, spectral skewness, spectral kurtosis, spectral slope, spectral drop, spectral roll-off, and so on. The amplitude spectrum can be described as:

$$A[k] = \left| F[x(n)] \right| \tag{11}$$

where $F[\cdot]$ refers to the discrete Fourier transform, $|\cdot|$ refers to the mode-taking operation, $A[\cdot]$ refers to the signal amplitude spectrum, and k is the frequency point number.

1) Spectral center of mass. The spectral center of mass (SC) is the frequency-weighted average of the amplitude spectrum, also called the center of gravity of the amplitude spectrum. The specific calculation formula is as follows:

$$f_{SC} = \sum_{k=0}^{N-1} k\,p[k] \tag{12}$$

where k is the frequency number and $p[k]$ is the normalized amplitude spectrum, which is calculated as follows:

$$p[k] = \frac{A[k]}{\sum_{k=0}^{N-1} A[k]} \tag{13}$$

The spectral center of mass is related to the frequency of the large magnitude and to the brightness of the sound. The higher the spectral center of mass, the brighter the sound.

2) Spectral Standard Deviation. The spectral standard deviation (SSp) describes the dispersion of the amplitude spectrum, and the specific formula is as follows:

$$f_{SSp} = \sqrt{\sum_{k=0}^{N-1} \left(k - f_{SC}\right)^2 p[k]} \tag{14}$$

In essence, it is the formula for calculating the standard deviation.

3) Spectral Skewness The spectral skewness (SSk) describes the asymmetric distribution of the amplitude spectrum on both sides of the spectral center of mass. The specific formula is as follows:

$$f_{SSk} = \frac{f_{moment3}}{f_{SSp}^3} \tag{15}$$

where $f_{moment3}$ is the third order moment, calculated as follows:

$$f_{moment3} = \sum_{k=0}^{N-1} \left(k - f_{sc}\right)^3 p[k] \tag{16}$$

With the difference in the sign of the spectral bias, the energy distribution shows different effects: when the spectral bias is zero, the energy distribution is the same on both sides of the spectral center of mass. When the spectral skewness is positive (positive bias), the energy on the left side of the spectral center of mass is higher than that on the right side, and the positive bias is intensified when its value increases, and vice versa.

4) Spectral Kurtosis. The spectral kurtosis (SK) describes the flatness of the amplitude spectrum near the spectral center of mass. The specific calculation formula is as follows:

$$f_{SK} = \frac{f_{moment4}}{f_{SSp}^4} \tag{17}$$

where $f_{moment4}$ is the fourth order moment, calculated as follows:

$$f_{moment4} = \sum_{k=0}^{N-1} \left(k - f_{SC}\right)^4 p[k] \tag{18}$$

With the difference in the value of the spectral peak state, the spectral distribution shows different effects: when the spectral peak state is 3, the distribution is normal. When the spectral kurtosis is less than 3, the distribution is flatter, and when its value decreases, the flatness increases. When the spectral kurtosis is greater than 3, the distribution is more sharp, and when its value increases, the sharpness increases.

5) Spectral Slope. The spectral slope (SSI) describes the rate of decline of the amplitude spectrum. The specific calculation formula is as follows:

$$f_{SS1} = \frac{1}{\sum_{k=0}^{N-1} A[k]} \frac{N \sum_{k=0}^{N-1} f_k A[k] - \sum_{k=0}^{N-1} f_k \sum_{k=0}^{N-1} A[k]}{N \sum_{k=0}^{N-1} f_k^2 - \left(\sum_{k=0}^{N-1} f_k \right)^2} \tag{19}$$

where f_k is the frequency corresponding to frequency point k.

6) Spectral Descent. The spectral descent (SD) also describes the decrease of the amplitude spectrum. The formula is as follows:

$$f_{\mathrm{sD}} = \frac{1}{\sum_{k=2}^{N-1} A[k]^{k-2}} \sum^{N-1} \frac{A[k] - A[1]}{k-1} \tag{20}$$

This feature is derived from perception studies and has been shown to correlate with human perception.

7) Spectral Roll-off. Spectral roll-off (SRO) refers to a frequency value below which 95% of the energy of the amplitude spectrum is concentrated. The specific formula is shown below:

$$\sum_{k=0}^{f_k \le f_{SEO}} A[k] = 0.95 \sum_{k=0}^{N-1} A[k] \tag{21}$$

The spectral roll-off is related to the cut-off frequency of the harmonics or noise.

2.2.3. Cepstrum characteristics. Given that the sound is a convolution of the impulse response of the excitation source and the filter unit, deconvolution is an unavoidable problem. Deconvolution algorithms can be divided into two categories: the first is “parametric deconvolution”, i.e. linear predictive analysis. The second category is “nonparametric deconvolution”, i.e. homomorphic analysis, also known as cepstrum analysis. The cepstrum features include Mel frequency cepstrum, Gammatone cepstrum, linear prediction cepstrum, etc. The cepstrum features are mainly used in the analysis of the cepstrum.

1) Mel frequency cepstrum coefficient. Different frequency components of a musical tone are perceived at different locations in the basilar membrane of the cochlea. The perception of sound frequencies by the basilar membrane of the human ear allows the division of the basilar membrane into a number of very small sections, each corresponding to a frequency group. These frequency groups can be designed as corresponding filters. For low-frequency components, the cochlea has a finer perception and a narrower filter bandwidth. For high-frequency components, the cochlea has limited perception and the filter bandwidth is wider. Although the filter bandwidth is nonlinear at linear frequencies, it is equally wide at Mel frequencies. Therefore, a cochlear-like function can be realized based on a delta filter bank on the Mel scale.The Mel frequency is the scale of pitch perception, which shows an overall linear relationship with linear frequencies in the range of 0-1000 Hz, as well as a logarithmic relationship above 1000 Hz. The conversion from linear to Mel frequencies is shown

in equation (22):

$$m = 2595\log_{10}\left(1+\frac{f}{700}\right) \tag{22}$$

where f is the linear frequency and m is the Mel frequency. The conversion of Mel frequency to linear frequency is shown in equation (23):

$$f = 700\left(10^{m/2595}-1\right) \tag{23}$$

MFCC is the discrete cosine transform (DCT) of the logarithmic power spectrum of a nonlinear Mel frequency, which is an alternative description of the power spectrum of a short-time signal [22]. In essence, it is a spectrum of the nonlinear compressed power spectrum, i.e., the cepstrum.

The process of calculating the Mel-frequency cepstrum coefficients is as follows:

Step1: Add a window to the signal in frames and take the magnitude spectrum by DFT transform for each frame.

Step2: Square the magnitude spectrum to get the power spectrum.

Step3: Design a triangular filter bank based on Mel scale.

Step4: Filter the power spectrum using the triangular filter bank and take the logarithm.

Step5: Perform DCT transform on the result obtained from Step4.

The MFCC obtained after the above five steps well combines the auditory characteristics of the human ear with the timbre formation mechanism of musical instruments, and is excellent in timbre expression. Since the single-frame MFCC only contains the static information within the frame, if the inter-song information is added, MECC differencing is required. The specific calculation formula is as follows:

$$d_t = \frac{\sum_{i=1} I\left(c_{l+i}-c_{l-i}\right)}{2\sum_{i=1}^{l} i^2} \tag{24}$$

where t is the frame number, i is the frame time delay, I is the order of the difference, c_t is the MFCC of frame t, and d_t is the time difference of frame t.

2) Gammatone cepstrum coefficient. The difference between GTCC and MFCC is only in the filter bank, GTCC uses Gammatone filter bank, the time domain expression of Gamatone filter is:

$$g(t) = b^n t^{n-1} e^{-2\pi bt}\cos\left(2\pi f_c t+\phi\right)u(t) \tag{25}$$

where f_c is the center frequency of the filter, $u(t)$ is the unit step function, n is the order of the filter, and parameter b is related to the equivalent rectangular bandwidth (ERB) of the Gammatone filter, which is calculated as follows:

$$b\left(f_c\right) = 24.7+0.108 f_c \tag{26}$$

3) Linear prediction of cepstrum coefficients. Linear prediction means that the current samples of a signal can be approximated using a linear combination of past samples. A unique set of prediction coefficients (LPC) can be obtained in the sense of minimum mean square error [23]. It is equivalent to modeling the signal with all poles, i.e., the AR (Auto-Regression) model. The specific formula is expressed as:

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + G\mu(n) \tag{27}$$

where a_k is the full-polar model coefficient of order k, G is the gain constant, and p is the AR model order. Its frequency domain expression form is as follows:

$$H(z)=\frac{G}{1-\sum_{k=1}^{p}a_k z^{-k}} \tag{28}$$

The model coefficients are obtained by minimizing the mean square error, and the specific solution methods are autocorrelation method, covariance and lattice type method and so on. The optimal predictor can be constructed through the model coefficients, and its formula is expressed as follows:

$$x(n)=\sum_{k=1}^{p}a_k x(n-k)+G\mu(n) \tag{29}$$

The LPCC can be derived by taking the inverse Fourier transform of $\log|H(z)|$. The LPCC can be derived from the LPC using the following derivation formula:

$$\begin{cases} a_c[0]=a_1 \\ a_c[n]=\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)a_k a_c[n-k]+a_n, & 1<n\le p \\ a_c[n]=\sum_{k=1}^{p}\left(1-\frac{k}{n}\right)a_k a_c[n-k], & n>p \end{cases} \tag{30}$$

where a_c is the linear prediction cepstrum coefficient, $LPCC=\{a_c\}$, n are the LPCC orders.

2.3. Piano music signal timbre feature extraction method

2.3.1. Piano Music Signal Preprocessing. In musical instrument recognition, the training of the instrument model and the recognition of the instrument are based on the selected timbre-related feature parameters. In order to make the extracted features more effective, the piano music signal is first analyzed and processed. The piano music signal preprocessing process is shown in Fig. 1, which is divided into basic processes such as sampling and quantization, removing mute segments, pre-emphasis processing, and frame-windowing.

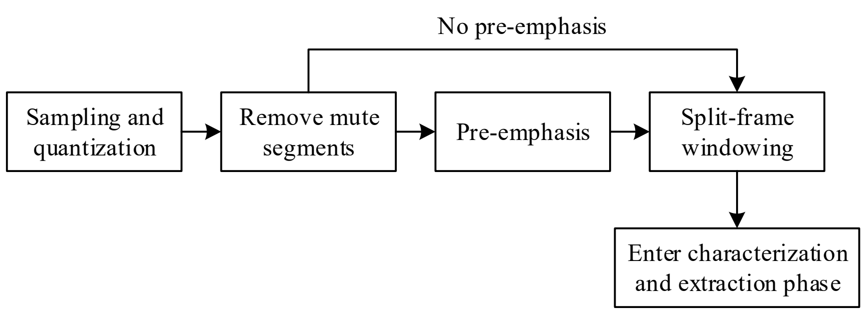


Fig. 1. Flow chart of piano sound signal preprocessing

1) Sampling and Quantization. Before the piano signal can be analyzed, the input sound must be digitized. Digitization of analog signals is generally achieved through sampling and quantization. According to the sampling theorem, when the sampling frequency is greater than twice the signal bandwidth, the original signal waveform can be reconstructed without distortion using an ideal filter. Each sample must be taken to quantize the sample points, and the more bits of quantization, the higher the accuracy, usually taking 8-bit or 16-bit digital quantities.

The sound of the piano in this experiment ends up being a piano tone pattern with a uniform sampling rate of 44.1kHz and a precision of 16bit, and the uniform format after sampling is Wav.

2) Removal of mute segments. Data preprocessing includes removing DC components, amplitude normalization and removing mute segments, this paper adopts the short-time energy-based double threshold endpoint detection to remove mute segments, the specific steps of preprocessing are as follows:

(1) Firstly, the signal is divided into frames to find out the short-time average energy, and then compare and judge based on the threshold value frame by frame.

(2) A rough judgment is made once based on a higher threshold α_1 selected on the short-time energy envelope of the speech, i.e., higher than this inter-value is the speech, and the music start and end points should be located outside the time points corresponding to the intersection of this threshold and the energy envelope.

(3) Determine a lower threshold α_2 on the average energy, and search from the previous intersection of the two ends to the left and right, respectively, to find the two points where the short-time energy intersects with the threshold α_2, which is the location of the starting and stopping points of the music segment determined by the double threshold method.

(4) Considering that there may be a minimum length indicating a pause between the notes of piano music signals, it is the end of the music segment that is judged to be less than the threshold α_2 to satisfy such a minimum length, which is actually equivalent to extending the length of the tailpiece.

3) Pre-emphasis. According to the principle of sound articulation, the amplitude of the high-frequency resonance peak is lower than that of the low-frequency resonance peak, and the higher the frequency, the smaller the spectral value. In order to improve the high-frequency resolution of the music signal, this paper carries out an overall spectral analysis of the entire frequency band, and pre-emphasis is realized by a first-order digital filter before the extraction of characteristic parameters. The transfer function of this filter is expressed as:

$$H(z) = 1 - \alpha z^{-1} \tag{31}$$

Where, α is the pre-emphasis factor, which is generally taken as a decimal close to 1. In this paper, we take 0.97. The pre-emphasis processing does the following transformations on the signal:

$$x(n) = x(n+1) - \alpha x(n) \tag{32}$$

The data obtained from the above equation is the pre-emphasized data, where $x(n)$ is the original audio signal.

4) Split-frame windowing. The feature extraction of piano music signal is based on the steady state signal. Therefore, before extracting the features of the piano music signal, frame segmentation processing is usually required. In order to ensure that the information is not lost, the frames need to overlap 1/3 to 1/2 frames between two frames, which is called frame shift. The theoretical formula for the number of frames in a piano music signal segment is:

$$N = \left[\frac{N_1 - N_0}{N_2 - N_0} \right] \tag{33}$$

where N denotes the number of frames, N_0 denotes the frame shift, N_1 denotes the total signal length, and N_2 denotes the frame length.

After all the music clips have been subframed, window operations are also required on the segmented frames in order to increase the continuity between frames, reduce edge effects, and minimize spectral leakage. Commonly used window functions in audio signal processing are rectangular window, Hanning window and Hamming window, defined as follows, M is the window length of the window function:

1) Rectangular window

$$w(n)=\begin{cases}1 & 0\le n<M\\ 0 & Other\end{cases} \tag{34}$$

2) Hanning Window

$$w(n)=\begin{cases}0.5\left(1-\cos(2\pi n/(M-1))\right) & 0\le n<M\\ 0 & Other\end{cases} \tag{35}$$

3) Hanming Window

$$w(n)=\begin{cases}0.54-0.46\cos\left(2\pi n/(M-1)\right) & 0\le n<M\\ 0 & Other\end{cases} \tag{36}$$

All three window functions have low-pass characteristics, and the main parameters that determine the nature and behavior of the time window in the frequency domain are the main flap, side flap, and roll-off characteristics, and the values of the main parameters of the different window functions are shown in Table 1.

Table 1. Main parameters of different window functions

Window Type	Main lobe width	First side lobe attenuation (dB)	Spectral roll-off (dB/ frequency doubling)
Rectangular window	4π/M	-13.3	20
Hanning window	8π/M	-32	60
Hamming window	8π/M	-43	20

The piano music signal has many prominent peaks and harmonics and is a composite signal. This makes the identification of peaks even more difficult because of the distortion caused by the improper selection of the time window among them. Therefore, in general, the time window is selected so that the amplitude of the side flaps is kept as low as possible, not only to alleviate the problem of spectral energy leakage, but also to make the peaks of the main flaps themselves more pronounced and prominent.

2.3.2. Tone Expression Spectrum Construction Based on Harmonic Structure:

1) Base Tone Detection. Autocorrelation function is a commonly used method for detecting base tones in the time domain. Autocorrelation analysis implies that if the signal is periodic, it repeats itself after a complete cycle [24]. For in terms of similarity, the corresponding signal is always similar when the time shift lies in an integer multiple of the period or fundamental frequency.

(1) Short-time autocorrelation function. Let the time series of the signal be $x(n)$, and the i nd frame of the speech signal is $x_i(l)$ after the windowed frame-splitting process, where the subscript i denotes the i th frame, and let the length of each frame be L. The short-time autocorrelation function of $x_i(l)$ is:

$$R_i(p)=\sum_{l=1}^{L-l}x_i(l)x_i(l+p) \tag{37}$$

where p is the amount of time delay.

The short-time autocorrelation function has the following important properties:

(a) If $x_i(l)$ is a periodic signal with period T, then $R_i(p)$ is also a periodic signal with the same period, i.e., there is $R_i(p)=R_i(p+\mathrm{T})$.

(b) The short-time autocorrelation function has a maximum value when $p=0$, i.e., the autocorrelation function of a periodic signal also reaches its maximum value when the amount of delay is $0,\pm T,\pm 2T,\cdots$.

(c) The short-time autocorrelation function is an even function, i.e., $R_i(p)=R_i(-p)$.

The main principle of the short-time autocorrelation function method of fundamental detection mostly utilizes these properties to determine the fundamental period by comparing the similarity between the original signal and its delayed signal. If the delay is equal to the fundamental period, the two signals have maximum similarity. Or it is straightforward to find the distance between the two maxima of the short-time autocorrelation function, which is the initial value of the fundamental period.

When using the autocorrelation function to detect the fundamental tone, a normalized autocorrelation function is commonly used with the expression:

$$ACF_i(p)=R_i(p)/R_i(0) \tag{38}$$

It has been pointed out in property (b) above that $R_i(0)$ is maximized at $p=0$, so the value of $ACF_i(p)$ is always less than or equal to 1.

(2) Base Tone Detection Based on Autocorrelation Function.The delay of the normalized autocorrelation function is the number of sample points, when the sampling frequency is f_s, the delay of each sample point is $1/f_s$. When using the correlation function method, the maximum value of the normalized autocorrelation function is found in the range between $T_{\min} \sim T_{\max}$, and the delay corresponding to the maximum value is the fundamental period. For piano music signals, it is common for note frequencies to lie between $27.5Hz \sim 4186.0Hz$, and $T_{\min}$ and $T_{\max}$ can be set accordingly.

In general, the larger the frame length N is, the more pronounced the peak of the autocorrelation function is, and the more data is available for analysis. When calculating the autocorrelation function ACF for long time signals, the longer the frame, the weaker the transient response. If the pitch shift is fast, using frames with large lengths will cause errors in the pitch detection results due to missing transient features. Therefore, there is a need to maintain a balance between the transient response and the accuracy of the fundamental frequency calculation.

The way to improve the transient response in the autocorrelation function based fundamental pitch detection algorithm is to dynamically change the frame length. The flow of base frequency detection calculation method based on autocorrelation function with variable frame length is shown in Fig. 2. The basic idea of the algorithm is to use the lowest fundamental frequency to determine the maximum frame length, and if the number of samples in the cycle obtained by the calculation is small, then it is no longer necessary to use this maximum frame length, and a shorter frame length is used.

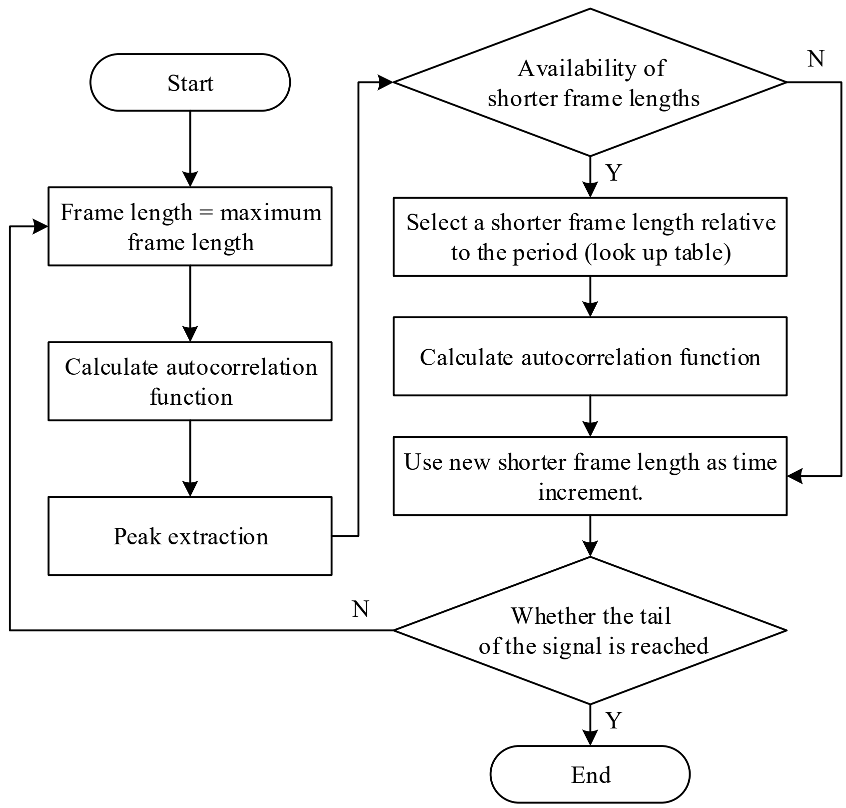


Fig. 2. Algorithm flow of adjusting frame length to improve transient

2) Definition and construction of timbre expression spectrum. Since the timbre of a musical instrument depends on the harmonic structure, similar musical instruments have similar harmonic structures, thus the harmonic structure can be defined as an objective index corresponding to the timbre of the instrument. Based on the existing discrete harmonic transform, the steps of piano timbre feature extraction based on harmonic structure are as follows:

Step1: Firstly, the music signal is sub-framed with a frame length of 0.5s and a frame shift of 0.25s, and no sub-framing is performed for the signal duration less than 0.5s.

Step2: According to the extraction method of harmonic structure, obtain the harmonic structure information of the piano music signal, normalize it and get the harmonic coefficients.

Step3: The piano tone expression spectrum is composed of discrete harmonic transform coefficients, first-order differential discrete harmonic transform coefficients and second-order differential discrete harmonic transform coefficients.

Through the above steps, all the piano timbre expression spectra constitute the piano timbre features based on the harmonic structure, which are used for the subsequent classification experiments, and the specific flow of the piano timbre expression spectrum construction algorithm is as follows:

Input: a frame of musical signal $x(n)$, the highest harmonic number of the constructed piano tone expression spectrum M, the fundamental period of the piano tone signal T_0 (optional).

Output: instrument timbre signature based on the piano timbre expression spectrum.

Step1: If the fundamental period of the piano tone signal is unknown, calculate the fundamental period of the piano tone signal based on the autocorrelation function method.

Step2: Make the window length $N = \lfloor T_0 \rfloor$ at the time of DHT transformation, and construct the harmonic structure information of the piano music signal according to Eq. (39), where $m = 1, \cdots, M$, and $M \le \left\lfloor \frac{T_0}{2} \right\rfloor$. That is:

$$DHT\left(m\right)=\frac{2}{N}\left|\sum_{n=0}^{N-1}x(n)w_{N}(n)e^{-j\frac{2\pi m}{N}n}\right| \tag{39}$$

Step3: Normalize $DHT\left(m\right)$ to get the piano timbre expression spectrum $DHTC\left(m\right)$.

Step4: Based on the piano timbre expression spectrum $DHTC\left(m\right)$, construct the first-order differential piano timbre expression spectrum $\Delta DHTC\left(m\right)$ and the third-order differential piano timbre expression spectrum $\Delta\left(\Delta DHTC(m)\right)$.

The overall process of piano timbre feature extraction based on harmonic structure is shown in Figure 3, which mainly includes five parts: data preprocessing, signal framing, fundamental detection, harmonic structure extraction and timbre expression spectrum feature construction.

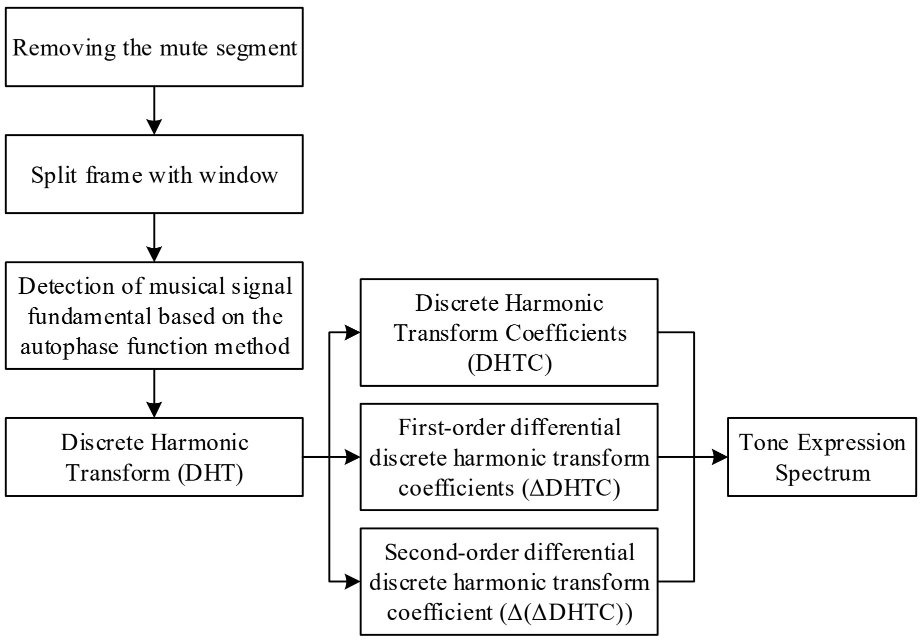


Fig. 3. Flow chart of tone feature extraction based on harmonic structure

2.3.3. Time-frequency cepstrum domain tone correlation based feature extraction. In order to analyze the features based on the piano timbre expression spectrum against those proposed in this paper, this section summarizes the common piano timbre-related features based on the video inverted spectral domain and their statistical features, and constructs a timbre feature extraction system for the piano musical sound signal.

The time-domain features include 26-dimensional features extracted directly from the piano music signal and 12-dimensional features extracted from the energy envelope of the piano music signal, totaling 38 dimensions. The frequency domain features include a total of 44-dimensional common audio descriptors extracted from the fast Fourier transform energy spectrum and power spectrum, and the mean and variance are obtained for all frames of the signal. The cepstrum domain features include 12-dimensional linear cepstrum prediction coefficients (LPCC) and 23-dimensional Meier frequency cepstrum coefficients (MFCC) and Meier difference cepstrum coefficients (ΔMFCC).

The extraction process of piano timbre-related features based on time-frequency cepstrum domain is shown in Fig. 4, and its specific steps are as follows:

Firstly, the piano tone signal is preprocessed, including removing the signal mute section, pre-emphasis, and calculating 10-dimensional global features for the overall energy envelope of the tone signal.

Secondly, the musical signal is divided into frames and windows are added, in which the window length is 0.5s and the window shift is 0.25s, and the autocorrelation coefficient, the over-zero rate and the RMS energy envelope are calculated for each frame of the signal, and eleven kinds of spectral features are calculated based on the STFT energy and power spectra, and the mean and variance of all the features obtained from all the frames are calculated.

Finally, the linear cepstrum prediction coefficients are computed based on the linear prediction coefficients, and the MFCC and first-order difference MFCC are computed based on the MFCC extraction principle for each frame of the signal, respectively.

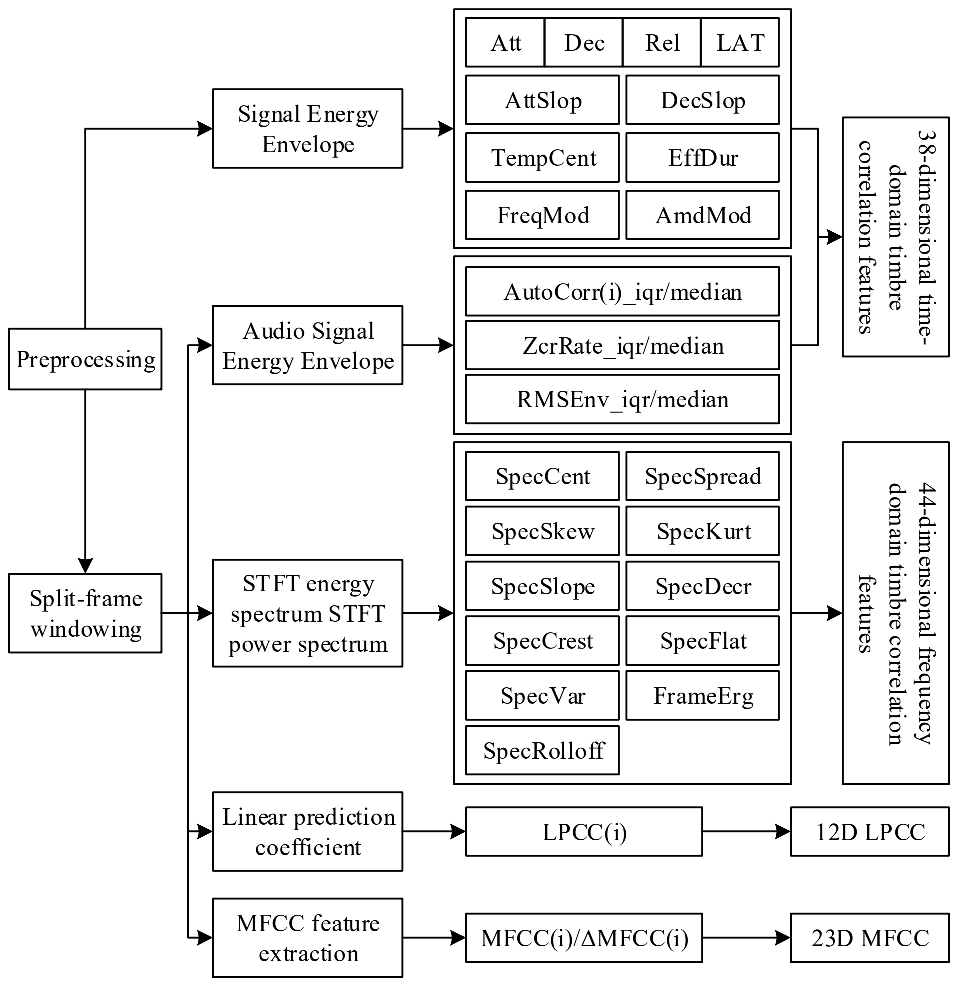


Fig. 4. Flow chart of timbre related feature extraction

2.4. Piano Teaching in the Context of Artificial Intelligence

The complexity and systematic nature of piano teaching is more obvious, requiring teachers and students to synergize their physiological, psychological, musical sense and many other aspects. Systematic training can better stimulate students' understanding and feeling of music, make students' limbs more coordinated, playing thinking more flexible, and finally express music well. In order to clarify the change and influence of AI interaction technology on piano tone teaching, the way of application of AI interaction technology in piano tone teaching session should be considered first.

2.4.1. Basic teaching aspects of piano. Piano playing must be preceded by a solid foundation of technique. Basic training includes hand shape, sitting posture, key touching method, force generation method, rhythm, music reading and other parts, which requires close cooperation between students and teachers. Many piano teachers have taken various ways to mobilize students' interest, hoping to achieve the goal of happy teaching, but the results are often not very satisfactory. Intelligent Piano is able to carry out this work very well, bringing unexpected results and fun. For example, in the primary stage of teaching, after the teacher has finished his/her teaching duties, it is impossible for him/her to spare a lot of time to practice with the students, so a more effective way is to hire an accompanist. However, the ability and time of the chaperone cannot be guaranteed. The advantage of Smart Piano is that it can fulfill the accompanying work without the teacher's presence. The smart piano can communicate on the 4K display, demonstrate music and playing skills, and also allow students to learn together with animated images, realizing the synchronization of teaching, practicing, playing and sharing, so that students can learn through playing. Therefore, the intelligent piano is fully capable of doing the basic practicing work well. Sometimes, the effect of AI Artificial Intelligence Piano will even be better than the effect of teacher teaching. Learners learn from easy to difficult with the help of the pass form, and if a step is wrong, an AI teacher will correct it.

2.4.2. Piano teachers and AI AI pianos work well together. In the traditional piano teaching, generally use a teacher with a student "one to one" teaching method, while the intelligent piano is based on the collective class form, how to realize the transition between the two articulation, has become a very important work. This paper believes that the following two points should be done:

Firstly, teaching in combination with students' personal development. In the process of teaching, piano teachers can add various contents, including audio, video and pictures, according to the teaching objectives and the characteristics of the teaching subjects, so as to make the piano teaching more efficient and reasonable. For example, before practicing Paganini Etudes, the movie Paganini, the Magic King of Violin can be played to the students. This can make the learners understand the background of the creation of the work more deeply, as well as the process of the transformation of this work from a violin to a piano work.

Second, AI AI pianos can always work efficiently and with high quality. The potential for scaling that AI has allows it to work as a chaperone for long periods of time without having to worry about not being guaranteed chaperone time. During practice, if there are mistakes, the AI piano can also correct them until the player is able to play smoothly. The AI piano is also able to record the content of the lesson, so that students can review it at any time. The student information collected by the AI piano also enables the teacher to understand the student's learning situation more accurately.

3. Piano music sound simulation experiment results and analysis

According to the analysis of the basic theory of piano string vibration and other timbre synthesis and related timbre features and the discussion of the timbre feature extraction method of the musical signal, this paper adopts MATLAB to design a digital piano timbre recognition and intelligent synthesis system to realize the piano timbre simulation, in order to use it for piano timbre teaching. In order to verify the effectiveness of the system, this chapter tests the system and verifies the simulation effect through piano tone simulation experiments.

3.1. System testing and analysis of its results

3.1.1. System implementation. In order to perfectly complete the intelligent synthesis of piano tone recognition, the envelope function needs to be extracted, the envelope function can present the relationship between the change of piano tone intensity and time, the application of the envelope function can ensure that the process of intelligent synthesis is consistent with the change of the function to analyze the amplitude of international standard piano tone, and the waveform diagram of the piano tone is shown in Figure 5.

Analyzing the waveform diagram of the piano tone, it can be seen that under the premise of guaranteeing the time, it is necessary to accurately allocate the piano vibration energy, and the function expression is:

$$f(x)=\frac{x^2}{e^x} \tag{40}$$

Where x represents time and e represents amplitude, the envelope function can pave the way for intelligent synthesis of piano music recognition. Combined with the above analysis, the piano timbre and other spectral items are compared to learn that the different amplitudes of the instrument also change, through the timbre feature matrix for the extraction of the piano timbre and the envelope function for the analysis of the piano amplitude, and ultimately realize the piano timbre recognition and intelligent synthesis.

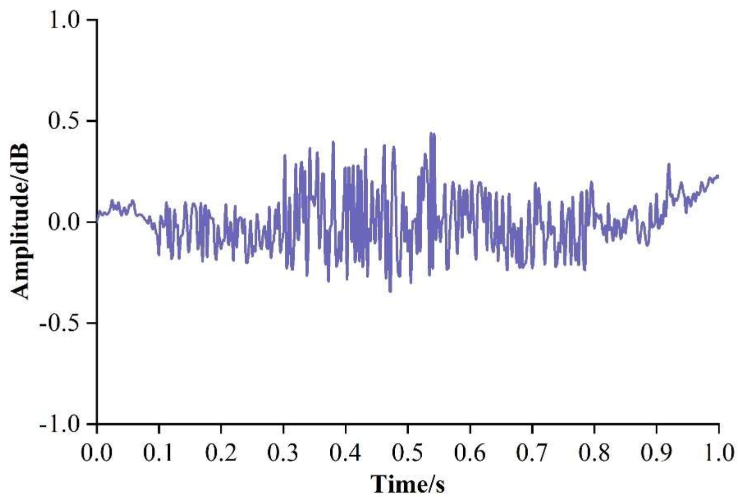


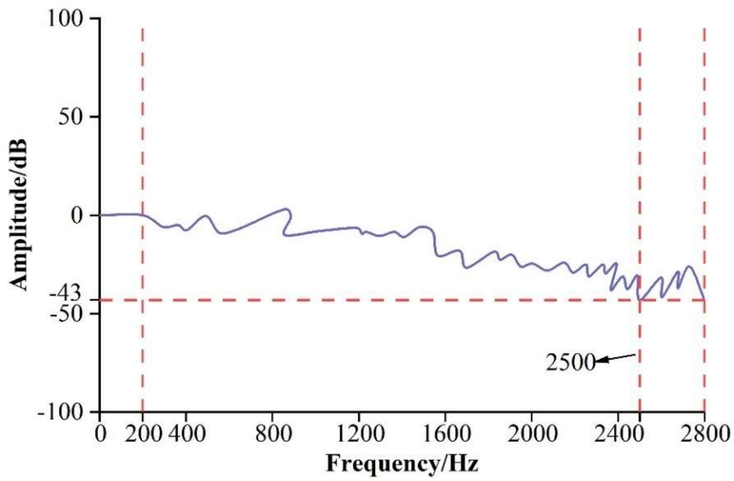
Fig. 5. Waveform diagram of piano timbre

3.1.2. System performance testing. The Piano Tone Recognition and Intelligent Synthesis System is supported by PC software and hardware, supporting Windows11 and WindowsXP operating systems, with a CPU requirement of 2GHz and a system disk space of 900 M. In the course of the test, the Piano Tone Recognition and Intelligent Synthesis System was installed and tested using 30 computers. The performance of the system and the correctness of the recognition were tested using computer technology.

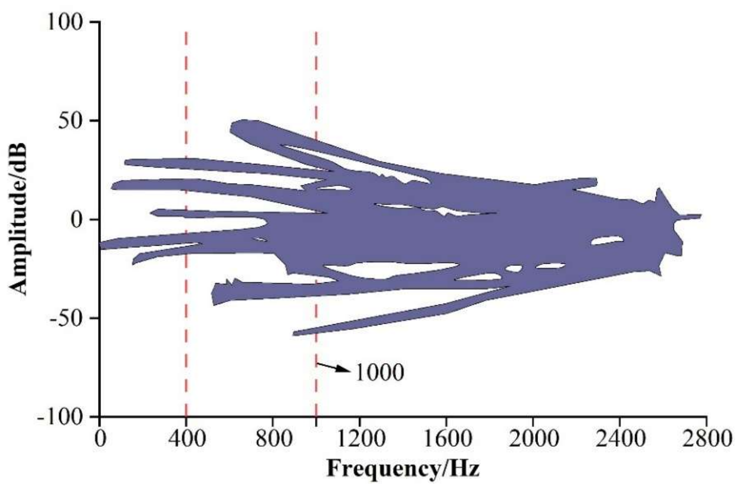
Through the opening and new construction of the piano sheet music, analyzing the music format and timbre test results, the opening of the piano sheet music only supports the normal opening of the current opening mode, the notes can only be added in the piano sheet music according to the correct format, and the sound and speed of the piano music playing is fixed. It is not possible to add dots in the newly created piano sheet music, and when deleting the notes at the end of the piano, it is necessary to recombine the notes of the piano sheet music, and finally change the file name of the piano sheet music and save it.

When adding, deleting and saving the notes of the piano in the above rules, the final test results of the piano timbre are qualified, indicating that the test results of the piano timbre are better. On this basis, the performance of the piano timbre recognition and intelligent synthesis system is tested by using MATLAB language, and the experimental results of the piano music signal timbre feature extraction method proposed in this paper and the comparison method are shown in Fig. 6's (a)~(b).

In the proposed method in this paper, when the frequency is 200 Hz, the corresponding amplitude decreases from 0. The amplitude of the piano tends to be stabilized when the frequency is 1600 Hz to 2400 Hz, without the interference of noise, and it has the lowest amplitude of -43 dB when the frequency is 2500 Hz and 2800 Hz. In contrast, in the comparison method, when the frequency is 300Hz~900Hz, the corresponding piano amplitude is higher, and with the change of frequency, the change range of the piano tone amplitude is larger, and there is the interference of noise. It can be seen that, as there is no noise in the proposed method in this paper, and the change of amplitude is clear and unambiguous, the clearer the change of amplitude is, the more clearly the piano tone can be recognized, which indicates that the piano tone recognition performance of the proposed method is better.



(a) Method mentioned in this article



(b) Comparison method

Fig. 6. System performance test comparison

On the basis of the analysis of piano timbre recognition performance, in order to test the correct rate of piano timbre recognition and intelligent synthesis system for timbre recognition, a piano song is selected for the experiment, the length of the experiment is 5 minutes, the correct rate of the recognition experiment is compared as shown in Fig. 7, and the shaded portion indicates the degree of improvement of the correct rate of this paper's method relative to the comparative method.

Analyzing Fig. 7, it can be seen that in the proposed method of this paper, when the playing time of the piano song is 1 minute and 2 minutes, the correct rate of the tone recognition is 51.07% and 59.42%, respectively; when the playing time is 3 minutes to 4 minutes, the corresponding correct rate of the tone recognition is about 79%, and near the end of the playing time, the correct rate of the tone recognition reaches 87.83%, which is higher than that of the start by 36.76%. In the comparison method, when the playing time of the piano song is 1 minute, the correct rate of tone recognition is 21.29%, when the time is 2, 3 and 4 minutes respectively, the correct rate of piano tone recognition is

39%~40%, which is in a relatively stable state, and when the playing time of the piano song goes to 5 minutes, the corresponding correct rate of tone recognition is 58.54%, which is higher than that of the beginning of playing by 37.25%, and the overall correct rate is between 21% and 59%. Comparison shows that the correct rate of the overall piano tone recognition of the proposed method in this paper is more than 51%, and the highest correct rate is about 88%, which indicates that the correct rate of piano tone recognition in the proposed method is higher, and the higher the correct rate of tone recognition the more perfect the system design is.

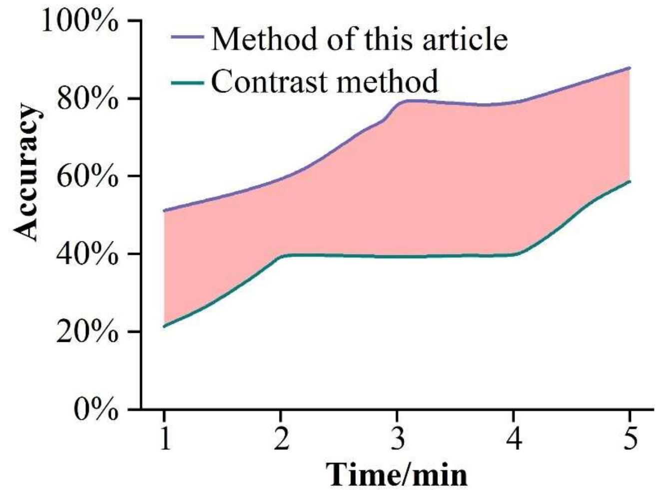


Fig. 7. The correct rate of piano timbre recognition

Combined with the above experiments, it is verified that the research-designed artificial intelligence-based piano timbre recognition and intelligent synthesis system has better performance and higher correct rate of timbre recognition, and can be applied to piano timbre teaching.

3.2. Piano music sound simulation experiment results and analysis

3.2.1. Piano Music Simulation Experiments. In this section, the piano piece "Two Tigers" is taken as an example, and the piano tone simulation experiment is carried out using the built piano tone recognition and intelligent synthesis system. The time-domain waveforms of the piano music sound simulation are shown in Fig. 8, where (a) is the music sound simulated by exponential envelope, (b) is the music sound simulated by folded linear envelope, (c) is the music sound simulated by segmented envelope curve, and (d) is the music sound simulated by the segmented envelope curve and filter modification at the same time. The effects of each simulation are analyzed as follows:

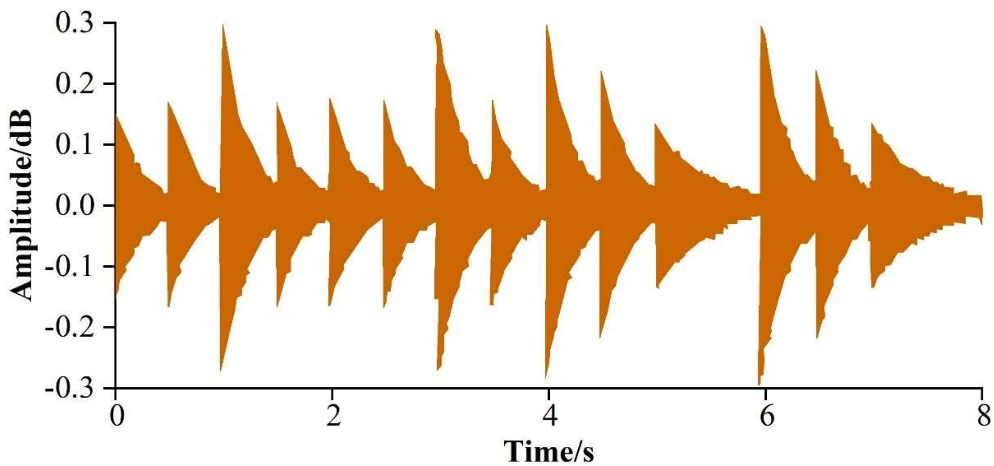
1) It can be heard from the music played that the exponential decay technique can well eliminate the transition noise between notes, and the decay speed determined according to the playing time makes the music sound more undulating, lively and powerful. However, through the audition judgment, it is found that there is still a slight murmur between the adjacent notes due to the sudden phase change, and the transition between the two notes is not very smooth.

2) The application of the folded envelope technique ensures that the amplitude of the previous note vibration decreases to 0 before the vibration of the next note starts, which effectively eliminates the noise between the note transitions. However, in the process of playing music, careful music professionals found that although the folded envelope method can make the transition between notes smooth, but the simulation of the music sounds relatively hard, can not present the softness of the piano music has a sense of softness, and the amplitude of change does not accord with the phenomenon of the stringed instruments.

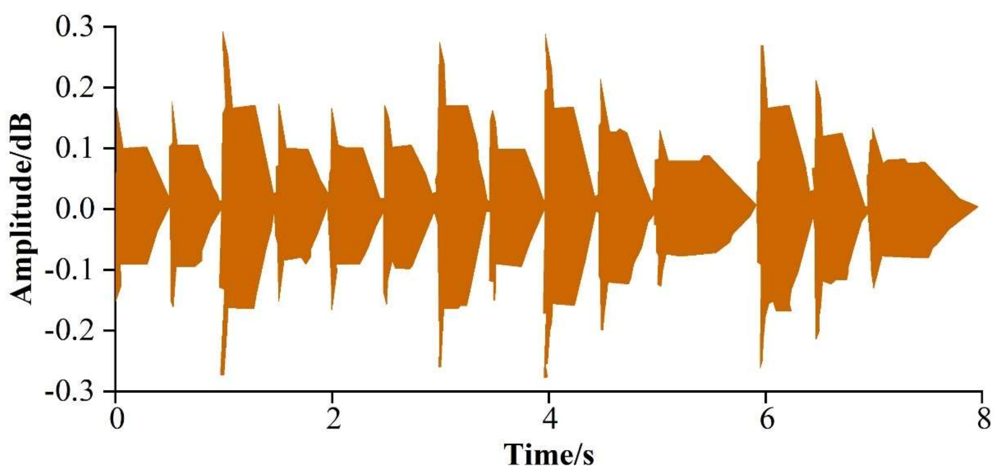
3) From the newly designed time-frequency inverse spectral domain envelope curve role of the map,

the envelope of the two-number curve changes gently, the start of the sound stage in a very short period of time will be able to increase the vibration amplitude to the maximum, to ensure that the piano's high efficiency of articulation, with the prolongation of time, the amplitude of the amplitude gradually attenuate when the note articulation is over, amplitude attenuation to 0, for the next note of the articulation of the preparation, better to avoid the transition of the transition between the neighboring notes produce Noise. Through audition, we found that the segmented envelope function simulation of the music effect is significantly better than the exponential envelope and folded envelope, the algorithm not only makes the transition between the notes smoothly, each tone changes softly on its own, completely eliminating the murmur, but also shows the piano music from the strings - struck by the articulation of the piano to the disappearance of the sound intensity of the soft change of the whole process.

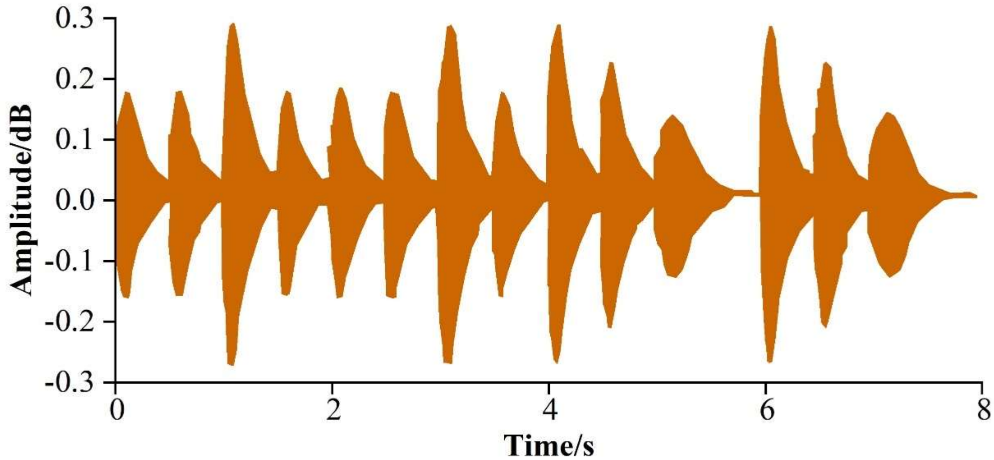
4) From the effect diagrams of the time-frequency cepstrum domain envelope function and the filter modeled by the spectral envelope after simultaneous modification, we know that this method not only harmonizes the articulation between the notes of the music, but also enhances the sound intensity of some signals. The sound is very close to that of a real piano with its orderly ups and downs.



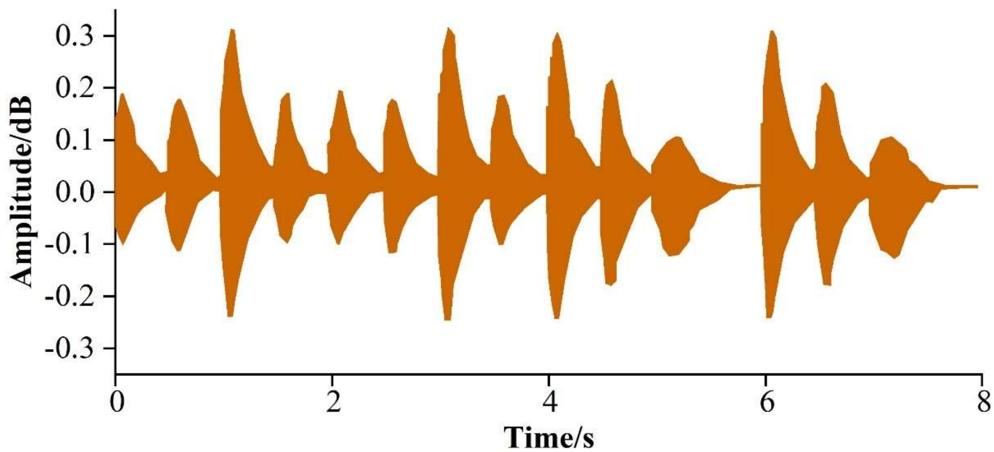
(a) Musical simulation using index envelope



(b) Musical simulation using polyline envelope



(c) Musical simulation using time-frequency inverted spectral domain envelope



(d) Music signal by frequency domain filtering processing

Fig. 8. Music simulation time waveforms

3.2.2 Experimental evaluation. In order to verify the effectiveness of the algorithm in this paper, this subsection will further evaluate the musical notes of the piano simulation from the following two aspects. One is to determine the readiness rate by detecting the fundamental frequency of the musical tone. The second is to compare the gap between the resonance peak frequency of the simulated musical sound and the resonance peak of the real piano music.

1) Base note detection experiment. First of all, the short-time autocorrelation function method is used to detect the fundamental frequency of the simulated musical signal of the piano, and the detection results are compared with the theoretical value of the fundamental frequency of each note in the sketch to obtain the accuracy. The detection results are shown in Table 2. For simplicity, only some note fundamental frequencies are shown in the table.

From the detected data, it can be seen that the musical signal simulated by the method of this paper is very close to the theoretical value of each note of the real piano instrument collected, and the accuracy rate reaches more than 99%, which proves the accuracy of the simulation of the piano sound.

Table 2. Test results

Notes in the simplified chart	Theoretical value/Hz	Detection value/Hz	Accuracy rate/%
1	272.54	272.97	99.84%
2	295.44	295.13	99.90%
3	330.52	331.08	99.83%
1	264.57	264.29	99.89%
3	330.58	332.47	99.43%
4	350.19	353.64	99.02%
5	394.28	395.81	99.61%

2) Estimation of resonance peak. The resonance system of a musical instrument is an important factor in determining its timbre, which is mainly reflected in the way of resonance to different frequency responses, which is reflected in the sound as a rise in the energy of certain frequency bands in the spectral structure. In the previous section, we have extracted the spectral envelope of the collected piano monophonic signal and calculated the frequency of the resonance peaks. Therefore, the same method can be used to extract the spectral envelope of the music signal simulated in this paper, and estimate its resonance peak point, in order to compare the simulation results with the real piano music value.

The schematic representation of the spectral envelope extraction of the musical signal simulated in this paper is shown in Fig. 9. The shape of the presented spectral envelope curve is very similar to the real piano music.

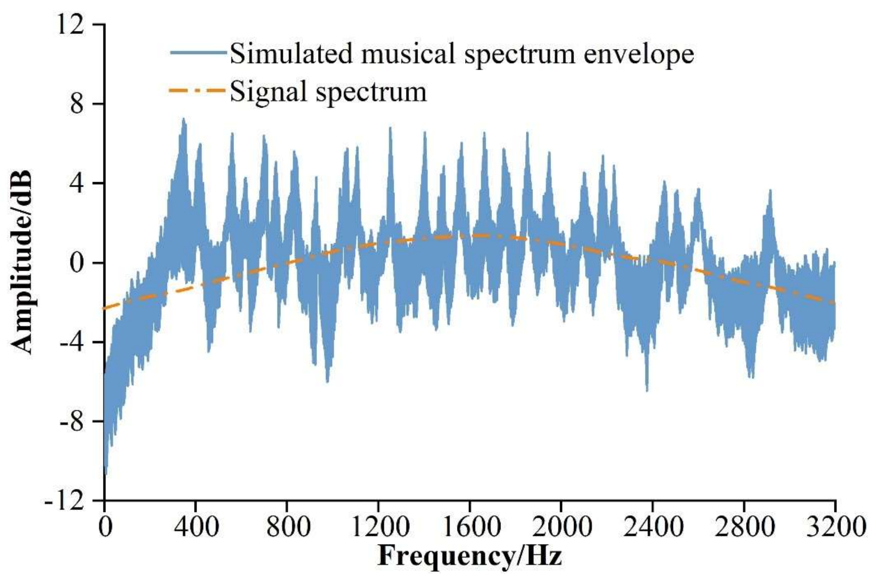


Fig. 9. The spectrum envelope diagram of the simulated music

In order to make the experiment more convincing, the values of the first resonance peak point, the second resonance peak point, and the third resonance peak point are calculated to be 1489.34 Hz, 4975.38 Hz, and 9013.86 Hz, respectively, and the difference between each resonance peak point and the real value is not big, which is in the range of resonance generated by the piano resonance system, and thus proves once again that this paper's time-frequency inverted spectral envelope modeling of the filter design of the effectiveness of the filter design modeled by time-frequency inverted spectral envelope in this paper.

3.3. Evaluation of the effectiveness of teaching piano tone

In order to verify the effect and application performance of the proposed piano music signal timbre feature extraction method in realizing the combination of AI interaction and piano timbre teaching concept, this paper analyzes through experiments and obtains the time-domain waveforms of the data analysis samples for quantitative assessment of the effect of the combination of AI interaction and piano timbre teaching concept, as shown in Fig. 10.

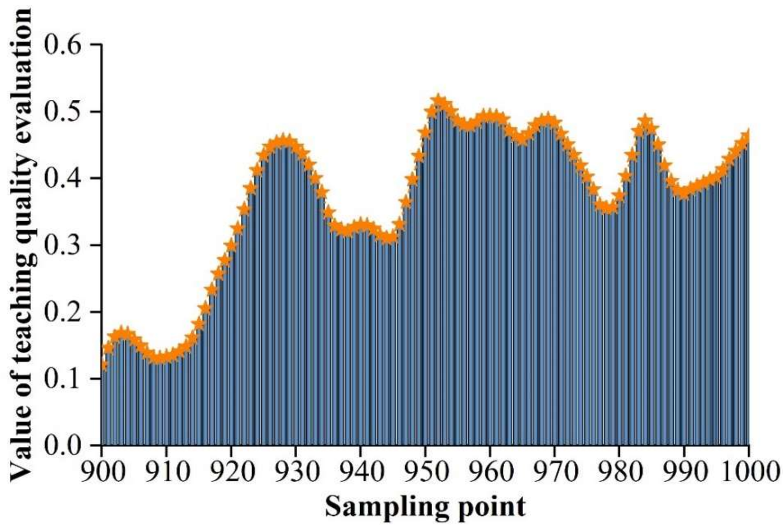


Fig. 10. Time domain distribution of quantitative analysis of teaching effect

Taking the data in Figure 10 as a test sample, the confidence level of teaching quality evaluation of the combination of artificial intelligence interaction and piano timbre teaching concept is analyzed, and the analysis results are shown in Figure 11.

Analysis of Figure 11 shows that the method proposed in this paper can effectively promote the combination of artificial intelligence interaction technology and piano tone teaching concept, and the confidence level of quantitative regression analysis is high, reaching 0.428 when the number of sampling points is 600, indicating that the method in this paper can realize the piano tone teaching based on artificial intelligence interaction and achieve better teaching quality.

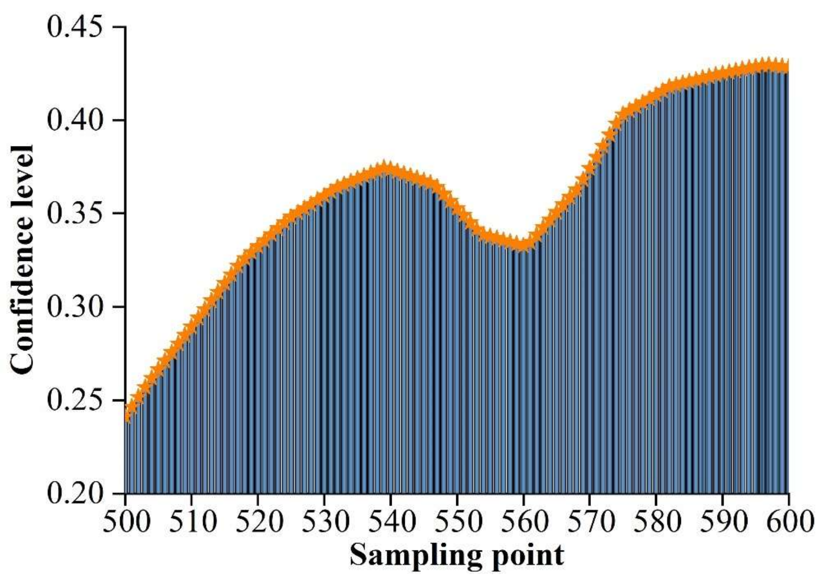


Fig. 11. Quantifying the confidence level of regression analysis

4. Conclusion

Based on the theory of timbre synthesis and timbre features, this paper proposes a timbre feature extraction method for piano musical signal, constructs a piano timbre recognition and intelligent synthesis system, and then realizes the construction of piano timbre teaching mode under the background of artificial intelligence interaction.

When utilizing the designed piano tone recognition and intelligent synthesis system, the corresponding amplitude decreases from 0 when the frequency is 200Hz, and the amplitude of the piano tends to be stable from 1600Hz to 2400Hz without the interference of noise, and reaches the lowest amplitude of -43dB at 2500Hz and 2800Hz. While using the comparison method, the corresponding piano amplitude is higher, the variation range of piano tone amplitude is larger, and there is the interference of noise. Meanwhile, in the 5-minute piano music simulation experiment, the correct rate of timbre recognition of this paper's method finally reaches 87.83%, which is much higher than that of the comparison method's 58.54% correct rate, which indicates that the piano timbre recognition performance of the proposed method in this paper is better.

In the piano sound simulation experiment, this paper takes the piano song “Two Tigers” as the research object, and its results show that: the method of this paper can not only make the musical notes of the notes of the articulation of the harmony, and some of the signals of the intensity of the sound has been enhanced. The music sounds orderly and close to the real piano. From the perspective of fundamental frequency detection and resonance peak frequency estimation, the music signal simulated by this method is very close to the theoretical value of each note of the real piano instrument, and the accuracy rate reaches more than 99%, which is highly accurate in the simulation of piano sound. The difference between each resonance peak and the real value is within the range of resonance generated by the piano resonance system, which proves the effectiveness of the filter design modeled by the time-frequency inverted spectral envelope in this paper.

In addition, in the evaluation of piano tone teaching effect, the method proposed in this paper can effectively promote the combination of artificial intelligence interaction technology and piano tone teaching concept, and the confidence level of quantitative regression analysis reaches 0.428 when the number of sampling points is 600, which indicates that the method in this paper is able to realize the high-quality teaching of piano tone under the background of artificial intelligence interaction.

References

[1] Reymore, L. (2022). Characterizing prototypical musical instrument timbres with Timbre Trait Profiles. Musicae Scientiae, 26(3), 648-674.

[2] McAdams, S., & Siedenburg, K. (2019). Perception and cognition of musical timbre. Foundations in music psychology: Theory and research, The MIT Press, 71-120.

[3] McAdams, S. (2019). Timbre as a structuring force in music. Timbre: Acoustics, perception, and cognition, 211-243.

[4] Arthurs, Y., Beeston, A. V., & Timmers, R. (2018). Perception of isolated chords: Examining frequency of occurrence, instrumental timbre, acoustic descriptors and musical training. Psychology of Music, 46(5), 662-681.

[5] Fischer, M., Soden, K., Thoret, E., Montrey, M., & McAdams, S. (2021). Instrument timbre enhances perceptual segregation in orchestral music. Music Perception: An Interdisciplinary Journal, 38(5), 473-498.

[6] Reymore, L., & Huron, D. (2018). Identifying the perceptual dimensions of musical instrument timbre. In Proceedings of the 15th International Conference on Music Perception and Cognition (pp. 372-377).

[7] Li, L. (2022). Design and Application of the Piano Teaching System Integrating Videos and Images. Scientific Programming.

[8] Yang, X. (2020). Research on auditory coordination in piano performance teaching. Arts Studies and

Criticism, 1(3).

[9] Wei, C. (2021). A Study of piano timbre teaching in the context of artificial intelligence interaction. Computational Intelligence and Neuroscience, 2021(1), 4920250.

[10] Phanichraksaphong, V., & Tsai, W. H. (2023). Automatic Assessment of Piano Performances Using Timbre and Pitch Features. Electronics, 12(8), 1791.

[11] Li, S., & Timmers, R. (2020). Exploring pianists' embodied concepts of piano timbre: An interview study. Journal of New Music Research, 49(5), 477-492.

[12] Zhang, L. X. (2019, June). The "four main factors theory" of piano teaching and its systematic thinking. In 2nd International Seminar on Education Research and Social Science (ISERSS 2019) (pp. 161-164). Atlantis Press.

[13] Qi, W., Dong, X., & Xue, X. (2021). The Pygmalion effect to piano teaching from the perspective of educational psychology. Frontiers in Psychology, 12, 690677.

[14] Zhang, D. (2019). Application of audio visual tuning detection software in piano tuning teaching. International Journal of Speech Technology, 22, 251-257.

[15] Stambaugh, L. A., & Nichols, B. E. (2020). The relationships among interval identification, pitch error detection, and stimulus timbre by preservice teachers. Journal of Research in Music Education, 67(4), 465-480.

[16] Dai, Y. (2023). Teaching Integration of Piano and Traditional Music Elements in Colleges and Universities Based on Network Flow Optimization. Applied Mathematics and Nonlinear Sciences, 9(1).

[17] Jiang, H. (2024). Machine learning in teaching piano touch technique and tonal expression. Journal of Electrical Systems, 20(4s), 404-415.

[18] Lv, J. (2017, July). Exploring the Relationship between Piano Playing Skills and Piano Sound. In 2017 International Conference on Sports, Arts, Education and Management Engineering (SAEME 2017) (pp. 336-339). Atlantis Press.

[19] Liu, L. (2022). TRAINING MEASURES OF PIANO TEACHING PERFORMANCE SKILLS FROM THE PERSPECTIVE OF COGNITIVE PSYCHOLOGY. Psychiatria Danubina, 34(suppl 5), 88-88.

[20] Wei, Y., Gan, L., & Huang, X. (2022). A review of research on the neurocognition for timbre perception. Frontiers in Psychology, 13, 869475.

[21] Wang, L. (2019). How to Coordinate Audio, Touch and Vision in Piano Performance Teaching. In 2019 1st International Education Technology and Research Conference (IETRC 2019) (pp. 762-765).

[22] Shoya Kawaguchi & Daichi Kitamura. (2023). Amplitude Spectrogram Prediction from Mel-Frequency Cepstrum Coefficients Using Deep Neural Networks. Journal of Signal Processing(6),207-211.

[23] Kethireddy Rashmi, Kadiri Sudarsana Reddy & Gangashetty Suryakanth V..(2022).Exploration of temporal dynamics of frequency domain linear prediction cepstral coefficients for dialect classification. Applied Acoustics.

[24] Tao Liu,Laixing Li, Yongbo Li & Khandaker Noman. (2024). Sliding time–frequency synchronous average based on autocorrelation function for extracting fault feature of bearings. Advanced Engineering Informatics(PD),102876-102876.