

## *Retraction*

# **Retracted: Pitch Evaluation of Matouqin Chamber Music Performance Based on Artificial Neural Network**

### **Mobile Information Systems**

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 2023

Copyright © 2023 Mobile Information Systems. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] G. Hu, "Pitch Evaluation of Matouqin Chamber Music Performance Based on Artificial Neural Network," *Mobile Information Systems*, vol. 2022, Article ID 7868975, 10 pages, 2022.

## Research Article

# Pitch Evaluation of Matouqin Chamber Music Performance Based on Artificial Neural Network

Guangying Hu 

*School of Art and Design the Inner Mongolia University of Science and Technology, Baotou 014010, Inner Mongolia, China*

Correspondence should be addressed to Guangying Hu; 2018933@imust.edu.cn

Received 23 April 2022; Revised 8 June 2022; Accepted 8 July 2022; Published 31 July 2022

Academic Editor: Yang Gao

Copyright © 2022 Guangying Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to study the pitch evaluation of Matouqin chamber music performance based on artificial neural network, this paper puts forward the relevant theories in the fields of human ear auditory perception system, auditory psychology, music theory knowledge, and pattern recognition. This paper extracts the auditory image features of chords and then establishes a sparse representation classifier model for chord recognition and classification. Scale-invariant feature transformation (SIFT) and spatial pyramid matching (SPM) are used to extract the detailed features of chord auditory images. The experimental results show that the highest correct recognition rate of the chord recognition algorithm based on the auditory image proposed in this paper is 76.2%, which is 20.4% higher than that of MFCC feature based on human auditory characteristics.

## 1. Introduction

The exploration of Matouqin chamber music performance technology can be traced back to the 1960s [1]. This is an era of vigorous development of Matouqin playing technology. Through the improvement of piano making technology, Chinese and Mongolian performers have developed treble Matouqin, alto Matouqin, sub alto Matouqin, bass Matouqin, and so on. It covers the range structure system required by chamber music orchestration. After half a century of continuous innovation and exploration, it has finally formed an independent Matouqin chamber music performance technology (different from Matouqin solo performance method). Due to the physical effect of sound transmission, the strength of the sound heard by human ears in the high range and low range is different under the same strength. The sound played in the high range is weak and the sound played in the low range is strong. This is because the amplitude and frequency of high range and low range in sound transmission are different. The sound amplitude played in high range is small and the frequency is high, and the sound amplitude played in low range is large and the frequency is low. When we appreciate symphonic works, we will find that the sound of 20 violins is less than that of

timpani [2]. Due to the differences in amplitude and frequency of instruments with different voices, the research on the orchestration of instruments with different voices began when the western national chamber music originated in the 14th century. Relying on the exploration of the volume structure, a standardized indoor and orchestration regulation has been formed. This standard indoor orchestra was initially applied to the court and gradually popularized in the mid-19th century. Its complete teaching and theoretical system of chamber music was formed and gradually improved in the 16th century, and a complete application system of performance technology has been formed in the 18th century. In the exploration of Matouqin chamber music, it is based on the string ensemble mode of Western chamber music, and integrates the singing timbre and volume of long key in short key for imitation exploration. Therefore, the strength application processing technology in Matouqin chamber music performance method is essentially different from Matouqin solo [3].

## 2. Literature Review

In recent decades, with the continuous development of information technology and the rapid popularization of

multimedia technology in daily life, more and more multimedia information has poured into the Internet, and digital music information, as an important component of digital multimedia information, also shows a rapid growth trend as shown in Figure 1. Sabir et al. and others found that the basis of sound production in chamber music performance method is ensemble. In the regulated orchestration, the representative instrument of each voice part will generally be selected, in which the general structure mode of high musical instrument, medium musical instrument, and bass instrument will inevitably appear. In this structure, composers often add subbaritone and bass instruments to the orchestration to enrich the texture structure of harmony and polyphony. In this century's Matouqin chamber music, there will also be the orchestration structure of the combination of Matouqin and wooden and copper groups in western music [4]. Meftah et al. and others believe that because of this, the volume control of different instruments in the performance of Matouqin chamber music will refer to the orchestration structure, and the volume effect after the superposition of ensemble volume needs to be considered in the joint performance [5]. Liu et al. and others first adopted the split volume processing method: the specific strength processing method of splitting into single parts in the overall harmony framework [6]. Monroe et al. and others take the Matouqin chamber music work "Chronicle of the wolf" as an example, which is a trio composed of the first Matouqin, the second Matouqin, and the piano. Since the piano has three parts: high, middle, and third, the first and second Matouqin are both the same part. The Matouqin (which covers the middle sound area and high sound area in the three octaves) plays two parts (high sound part and low sound part) of the piano in the first to twelfth bars of the introduction of the work [7]. Ahf et al. and others found that the main melody now starts with the triad of the low voice part, transitions to the high voice part in the fourth bar, transitions to the low voice part in the fifth bar, and is completed by two voice parts in the eighth bar. In this paragraph, the composer is marked with MP (medium weak), and the two parts are separated in volume processing. The main melody is played by MP (medium weak), and the accompaniment melody is played by P (weak). When the two parts are combined, the auditory effect of medium and weak volume is produced [8]. Wang et al. and others found that the ensemble structure of Matouqin and Qin starts from the 13th bar. The first Matouqin and the second Matouqin are a parallel main melody of polyphonic structure, and the sound part is a rhythmic accompaniment melody. The composer marks the volume prompt of P (weak) here. Due to the duet structure of Matouqin, in the volume processing of splitting in each sound part, each sound part can be played as P (weak), and the two sound parts are strong at the same time until the fifteenth bar is *f* (strong). In the strong position, the two sound parts are split into MF (medium strong) [9]. Carpenter and others believe that this strong and weak segmentation method stems from the different prominence of strong and weak tones. When two weak tones are superimposed, it produces a weak volume effect in the sound effect

[10]. However, Batool et al. and others believe that when two medium and strong volumes are combined, strong sound will produce strong sound effect, and strong sound will have a greater impact on sound in hearing. Just as in listening to symphony works, weak sound will produce synesthesia effect of distance feeling, and strong sound will produce sound impact feeling of close distance [11]. Therefore, in the cutting and integration of the strength treatment of chamber music performance, we need to first consider the constant volume of the ensemble of musical instruments, split each sound part on this basis, reasonably plan the strength of each sound part, and then integrate to form a complete constant volume structure. In the performance of chamber music, the music processing method of theme is the soul of the work. The theme structure is divided into presentation theme, dialogue theme, and imitation performance method in a solo paragraph. Schneider et al. and others found that in the paragraph of presentation theme, the presentation theme is generally in the highest position of the specified volume in the standard volume. Take the Matouqin chamber music work "Recollection (for Matouqin and Woodwind Group and brass group)" as an example [12]. Matveev et al. and others believe that the theme melody in the works is first played by the clarinet, and the other parts are long accompaniment or pause. The composer's musical expression is marked MF (medium weak) here. During the performance, the clarinet is played in MF (medium weak), while the string group is marked MP (medium weak), but it is completed by five parts together. Each part, after being divided into five parts, is played in PP (very weak). After the ensemble, it leads to the superposition of volume and presents the auditory effect of MP (medium weak) [13].

On the basis of this research, this paper proposes a study of intonation evaluation in Matouqin chamber music performance based on artificial neural network. Scale-invariant feature transformation (SIFT) and spatial pyramid matching (SPM) were performed on auditory images of different chords to extract detailed features of chord auditory images. The experimental results show that the method has certain development prospects. First, extract the auditory image features of the music chord segment, convert the one-dimensional music signal into two-dimensional image features, and then extract the local features of the auditory image, that is, the SIFT feature vector, and then use the SPM matching method to integrate the local feature vectors of the image into a feature vector to represent the features of the complete auditory image, that is, the chord features of music. Secondly, the pattern recognition method based on the sparse representation classifier (SRC) has achieved great success in image scene classification, object recognition, and face recognition. Subsequently, SRC has also been introduced into music genres, classical music classification, and music chord recognition. They have also achieved good results. Therefore, this paper uses the SRC method to identify chords. Finally, experiments are carried out under the optimal parameter settings, and the experimental results show that the recognition effect of the method using auditory image features and SRC recognition is the best.

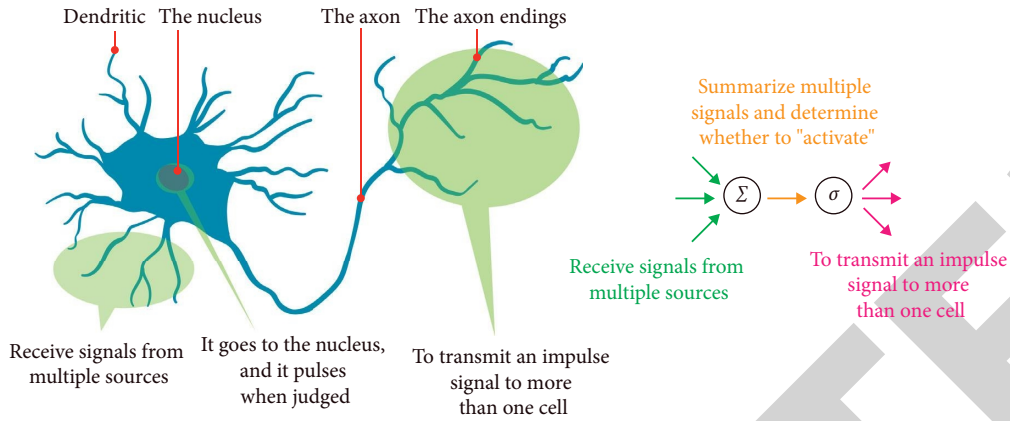


FIGURE 1: Abstraction of biological neuron structure by artificial neuron.

### 3. Method

Music is the product of the combination of science and art. Music recognition involves different disciplines such as physics, musicology, signal processing, art, and so on [14]. As the smallest component of the music signal, a chord can convey the harmony content, melody, rhythm, emotion, and other important information of music. As one of the important research topics in the field of music information retrieval, music chord recognition has a wide range of applications, such as music segmentation, similarity music retrieval, and humming retrieval. This paper introduces the generation mechanism of human voice, the basic attribute characteristics of music and human ear auditory system, so as to further deepen the research on music signal processing and the human ear auditory model. Sound is formed by the sound wave generated by the regular vibration of an elastic object [15]. The basic physical properties of sound include pitch, timbre, and intensity. These basic characteristics play a major role in the chord, rhythm, and melody of the middle and high levels of music. Therefore, the analysis of music signals requires researchers to know and master the basic theory of music. With the support of music theory, they can have a deeper research on music signals, so as to develop a better recognition algorithm. Pitch, that is, the height of sound is produced by the vibration of different objects. Its height is determined by the frequency of sound wave vibration. The two constitute a positive proportional relationship, that is, if the vibration frequency is high, the sound will be high. On the contrary, if the vibration frequency is low, the sound will be low. The vocal cord vibration frequency of a female singer is higher than that of a male singer, so the female voice is higher than the male voice. Human perception of pitch has a logarithmic relationship with the fundamental frequency as shown in equation (1). The unit of pitch is mel. For example, when the frequency of a sound signal is 1000 Hz, the pitch perceived by human ears is about 1000 mel.

$$T_{\text{Mel}} = 2595 \log\left(\frac{1+f}{700}\right). \quad (1)$$

The tuning curve of the piano is shown in Figure 2. Since the maximum gap between bass and treble can reach dozens

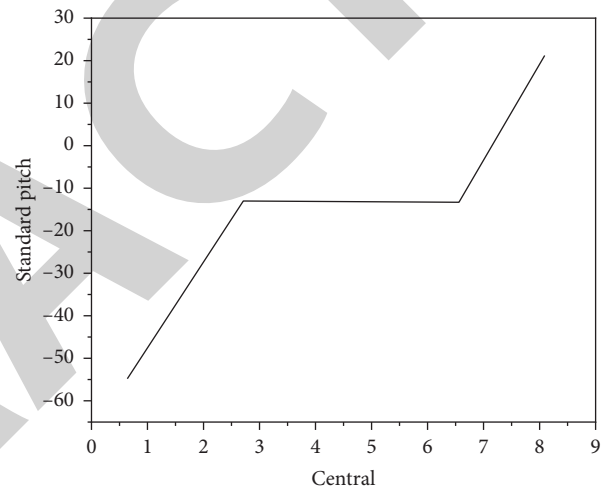


FIGURE 2: Tuning curve.

of sound points, if the pitch is determined according to the twelve average law when the piano is actually tuned, the bass area should be down and the treble area should be up, so as to produce a correct scale feeling [16]. International standard pitch refers to the frequency of note an above the central C of the piano, i.e.,  $a = 440$  Hz.

Rhythm refers to the absolute accurate height of all notes in the musical system and the relationship between them. It is a concept formed in the continuous development of music. There are three main categories of rhythm, namely, pure rhythm, five degree phase law, and twelve average law. Among them, twelve average law is a widely used representation of rhythm in the world. Twelve average law (hereinafter referred to as average law) is the most commonly used form of rhythm expression in western music [17]. It divides the adjacent tones in an octave into twelve semitones according to the principle of equal frequency ratio, in which the semitone represents the minimum distance value of the pitch in the twelve average law system. Generally, keyboard instruments adopt the average rhythm system, that is, the pitch difference of any two adjacent keys is halftone, and the frequency ratio is equal. See the following formula:

$$f_1 f_2 = 2^{1/12}. \quad (2)$$

Scale refers to each sound in the musical system. It is the distance unit between tones. In music theory, there are seven basic levels named A, B, C, D, E, F, and G (i.e., sound names), and these seven levels are made by the white keys of the piano. The piano has a total of 88 keys, so the above 7 sound level marks are recycled. The pitch of the same sound level in each group is different, and there is an “octave” difference between the same sound level in the adjacent two cycle groups. According to the law of the twelve mean law, the octave is divided into 12 equal parts, each of which is called semitone. Two semitones form a whole tone, in which the semitone is the smallest unit of music, and the whole tone and semitone form a double relationship in width [18]. Interval refers to the pitch distance between two levels, which is expressed in degrees. The number of tones refers to the sum of the number of semitones and whole tones contained between intervals. Degree refers to the number of sound levels contained between the root and crown sounds (i.e., the number of lines contained in the staff). On the staff, the interval relationship between two tones on the same line or at the same interval is called “one degree” or “same degree.” If one sound is on the line and the other sound is between the adjacent sounds, it is called “second degree.” The name of interval is determined by degree and tone number. Table 1 shows the naming rules of interval.

A chord is a group of three or more notes with a certain interval relationship. In other words, the simplest chord consists of three notes, and the complex chord can consist of five to seven notes. The most basic sound in a chord is called the root sound. Other sounds are divided into three degrees, five degrees, and seven degrees according to the distance between them. There are many kinds of chords. According to the number of notes in the chord, it can be divided into triad, seventh chord, ninth chord, etc. [19]. The triad contains three notes, namely, root, third degree, and fifth degree. It can also be divided into large and small triads and increasing and decreasing triads. A seventh degree is superimposed on the basis of a triad to form a seventh chord. Similarly, seven chords can also be divided into four types: large and small seven chords and increasing and decreasing seven chords. On the basis of seven chords, a ninth is superimposed to form nine chords. By analogy, we can get eleven, thirteen chords, and so on. Table 2 shows how different chords are named.

The process of hearing is: first, the sound wave is transmitted to the tympanic membrane through the external auditory canal, causing the vibration of the tympanic membrane, and then transmitted to the inner ear through the auditory ossicles, so that the receptors in the cochlea are stimulated to produce nerve impulses. Finally, the nerve impulse is transmitted to the auditory center of the cerebral cortex through the auditory nerve, so as to form hearing. The external ear is composed of pinna, external audit meatus, and eardrum. The auricle transmits the sound heard to the tympanic membrane through the external auditory canal, which causes mechanical vibration and converts the sound

TABLE 1: Interval naming rules.

Degrees	Interval name	Tone number
One degree	Pure degree	0
Two degrees	Minor second	0/2
	Major second	1
Three degrees	Minor third	1 1/2
	Minor third	2
Four degrees	Pure fourth	2 1/2
	Augmented fourth	3
Five degrees	Diminished fifth	3
	Perfect fifth	3 1/2

energy of sound waves into mechanical energy. Because the auricle is curly, it can better locate the high-frequency sound and directionally transmit the sound wave to the ear canal [20]. The external auditory canal is an approximately circular, uniform tube with one end closed. Its diameter is about 5 mm and its length is about 25 mm, which effectively protects the tympanic membrane from mechanical damage caused by external sound. Formant frequency of external auditory canal = sound velocity/sound wavelength. It is known that the sound wave length is about 4 times the length of the external auditory canal and the sound propagation rate is 340 m/s. Therefore, the formant frequency of the external auditory canal =  $340/(4 \times 0.025) = 3.4$  kHz, that is, the natural resonance frequency is about 3.4 kHz. The human ear is more sensitive to sound in some frequency ranges. The main reason is that the external auditory canal has resonance and diffraction effect on sound waves, resulting in high transmission gain of the external ear in the frequency range of 2 kHz ~ 4 kHz. Therefore, the main function of the external ear is to locate the sound source and amplify the sound. The tympanic membrane is located in the innermost part of the external auditory canal, separating the external ear from the inner ear, and the sound is transmitted to the inner ear by the vibration of the tympanic membrane [21].

The middle ear consists of tympanic membrane, tympanic chamber, three auditory ossicles, oval window, and circular window. The tympanic membrane is located between the outer ear and the inner ear, which plays a role in isolation. The utility model uses a circular window and an oval window to communicate with the inner ear and then establishes a connection with the outside world through the eustachian tube to balance the atmospheric pressure between the middle ear and the outside world. When the sound intensity is within a certain range, the auditory ossicles transmit sound waves in a linear form. However, when the sound intensity is very high, the auditory ossicles exhibit nonlinear propagation. The nonlinear propagation mode of auditory ossicles effectively protects the inner ear from mechanical damage. To sum up, the middle ear has two main functions: one is to amplify the sound pressure value on the tympanic membrane, and the other is to realize nonlinear transmission when the sound is very strong, so as to effectively protect the inner ear.

The inner ear is located in the deepest part of the skull and is composed of semicircular canal, oval window, and

TABLE 2: List of triads.

Name	An interval consisting of the third and root notes	An interval consisting of the fifth and the root	Chord name
Major chord	Major third	Perfect fifth	C, Cm, Cmaj
Minor triad	Minor third	Perfect fifth	Cm, Cmin
Increasing triad	Major third	Augmented fifth	Caug
Subtractive triad	Minor third	Diminished fifth	Cdim

cochlea. The semicircular canal and vestibular window are the direct receptors of the human body, which are related to the balance function of the body. The semicircular canal is composed of anterior semicircular canal, outer semicircular canal, and posterior semicircular canal. These three semicircular tubules are perpendicular to each other, similar to three-dimensional coordinate structure. The receptors located in the semicircular canal can feel the stimulation brought by rotating speed change. On the contrary, the receptor located in the vestibular window senses the movement of linear speed change [22]. Cochlea is the most important part of the inner ear and plays the greatest role in auditory perception. It is the receiver of hearing. The Basilar membrane is an important part of the cochlea. The Basilar membrane near the vestibular window is hard and narrow, while the part near the cochlear hole is soft and wide. The organ of Corti is located on the basement membrane and plays a sensing role. The potentials on both sides of the hair cell membrane on the organ of Corti change with the change of the fluid velocity in the cochlea. This change makes the auditory nerve release and inhibit. It is this change that converts the sound wave into nerve impulse and then completes the signal release process.

The masking effect is due to the frequency selectivity of the human ear to the sound, that is, when the strong sound and the weak sound exist at the same time, the strong sound is most easily detected by the human ear, while the weak sound is masked by the strong sound and difficult to detect. This phenomenon of increasing the hearing threshold of weak tone due to the existence of strong tone is called masking effect. The former is called masking sound and the latter is called masked sound as shown in Figure 3.

Whether a sound can be perceived by the human ear is determined by the frequency and intensity of the sound. The frequency range of sound that can be detected by ordinary human ears is 20 Hz ~ 20 kHz, and the sound intensity is -5 dB ~ 130 dB. The sound beyond this range cannot be detected by human ears. Within the normal hearing range, the most sensitive frequency band of human ear response to sound is 2 kHz ~ 4 kHz. Beyond this frequency band, the auditory sensitivity will be reduced. Hearing threshold refers to the value of the sound pressure level of the weakest sound that can be heard by human ears. The hearing threshold is related to the sound frequency function. The dotted line in Figure 3 represents the hearing threshold curve of human ear in a quiet environment. Human ear cannot hear the sound signal with sound pressure value lower than the hearing threshold. For example, when the sound pressure value of a pure tone signal is lower than the hearing threshold, the human ear cannot hear the sound signal [23]. In fact, the minimum value of human ear hearing threshold

is in the range of 3 kHz ~ 5 kHz, that is, the human ear is the most sensitive to weak signals in this frequency band. The hearing threshold outside this frequency band is much larger than that of this frequency band, that is, the human ear has poor sensitivity to the sound signal outside this frequency band. In the range of 0.8 kHz ~ 1.5 kHz, the threshold curve changes most gently, and the hearing threshold changes little with frequency [24].

If there is a strong sound signal, the listening threshold curve will change within its frequency range, that is, the listening threshold will be increased. This value is called masking threshold, as shown in Figure 3. In the neighborhood, the sound below the masking threshold is masked, so the human ear cannot hear the masked sound. Masking effect can be subdivided into simultaneous masking (also known as frequency domain masking) and isochronous masking (also known as time-domain effect). The difference between these two effects is whether the masking sound and the masked sound act at the same time. Isochronous masking can be further divided into front masking and rear masking. The former appears before the beginning of masking sound and the latter appears at the end of masking sound. Figure 4 shows three masking effects, with the horizontal axis representing the duration and the vertical axis representing the sound pressure level [25]. At the same time, masking occurs in the time period of 0 ~ 200 ms of masking sound, the front masking occurs in the first 20 ms of masking sound, and the rear masking occurs in 200 ms after masking sound disappears. As can be seen from Figure 4, the disappearance of isochronous masking is related to time. Generally, the duration of front masking is about 5 ~ 20 ms, while that of rear masking can last up to 50 ~ 200 ms.

Auditory image model (AIM) is a time-domain model, which simulates the auditory pathway according to the response state of human auditory system in different processing stages of sound signal, and then processes the signal effectively. "Auditory image" first appeared in the article published by Patterson in 1995. The model successfully formed the sound signals heard by human ears in the brain, and the initial consciousness was simulated as a neural representation. Aim provides a basic model for more researchers committed to audio research. The auditory image model is mainly composed of five basic functional modules b391, which are: (1) Transmitting the sound signal to the cochlear pre-processing (PCP) module of the oval window; (2) the module of the response of the cochlea to the basal membrane motion (BMM) of the signal; (3) neural activity pattern (NAP) in auditory nerve and cochlear nucleus; (4) strobe temporal integration (STI) module for generating auditory images; (5) form a stable auditory image (SAI) with auditory awareness, as shown in Figure 5.

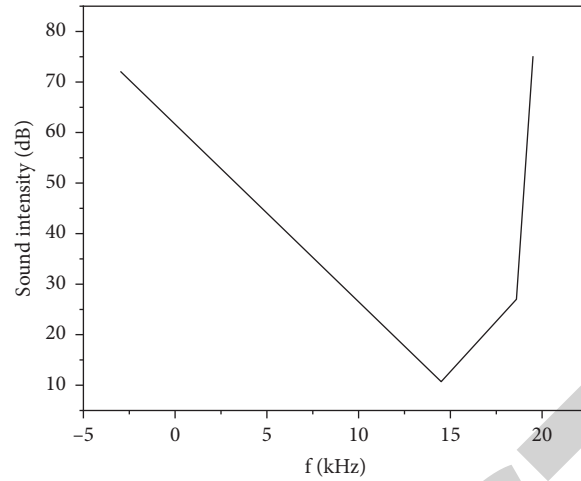


FIGURE 3: Masking effect.

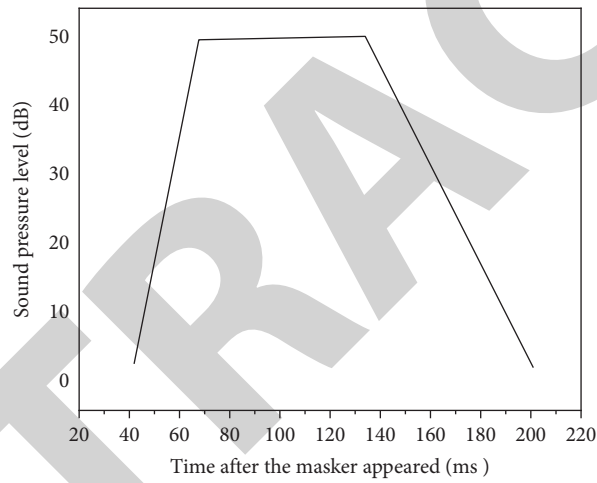


FIGURE 4: Intensity and duration of three masking phenomena.

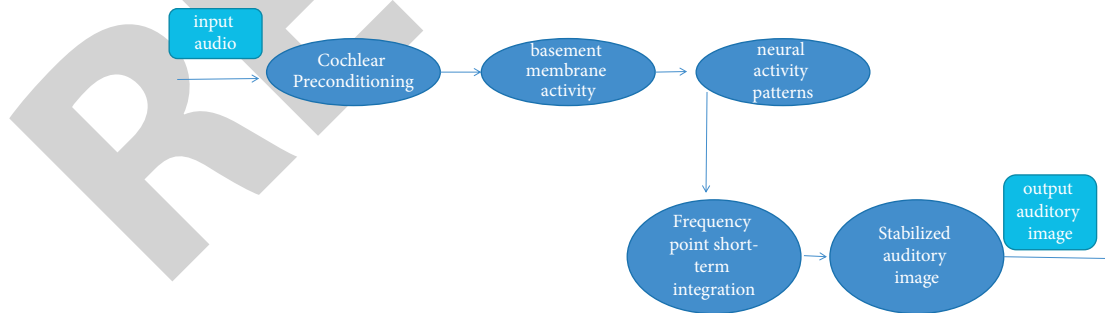


FIGURE 5: Auditory image structure.

AIM refers to the physiological structure and function of human ear and completes the simulation of human ear hearing through filter design. Each functional module forms a corresponding relationship with human ear hearing structure. The corresponding relationship between human ear structure, aim function blocks and implementation methods is shown in Table 3.

The sound frequency range that human ears can perceive is 20 Hz ~ 20 kHz. PCP module actually uses the filtering function of band-pass filter to simulate the response process of external ear and middle ear to sound signal. The signals beyond the hearing range of human ears will be filtered out, and the effective signals will be transmitted to the subsequent processing module for analysis and processing. Figure 6



TABLE 3: Corresponding relationship between human ear structure, AIM function blocks and implementation methods.

Human ear structure	Auricle and middle ear	Cochlea	Brain stem, nerve bed, cerebral cortex		
Auditory image module	PCP	BMM	NAP	STI	SAI
Implementation mode	Band-pass filter	Frequency analysis	Sharpening	Feature detection	Short term integration

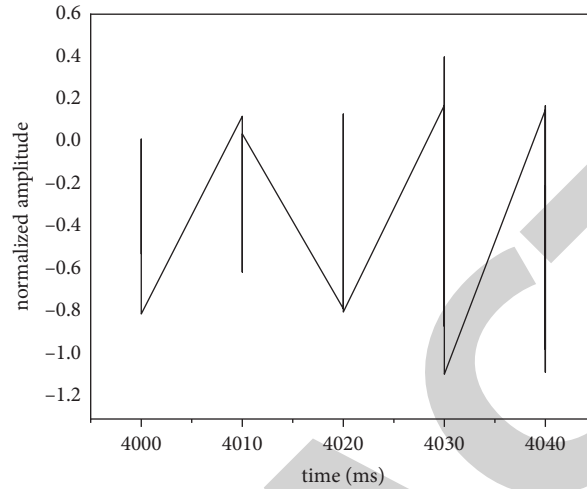


FIGURE 6: Waveforms before and after PCP processing.

shows the original audio of the large chord segment and the waveform after PCP processing. The upper figure is the original audio waveform, the lower figure is the waveform after PCP processing, the horizontal direction is the time axis, and the vertical direction is the normalized amplitude.

#### 4. Results and Analysis

According to the principle of parallel accumulation, structural music is the simplest structural method in musical forms, also known as the principle of parallel combination. It is characterized by the accumulation in the horizontal extension between the music forms with different degrees of contrast and renewal. Each part has the same scale and weight, and can express clear music content. There are two types of music forms in line with this principle which are illustrated below; when it comes to the basis of juxtaposition, when we start from the simplest musical form, it is its horizontal accumulation that can gradually form the structure of the principle of juxtaposition and combination. According to the theoretical interpretation of Professor Yang Ruhuai in the article “On marginal musical forms,” the structure of a musical form also conforms to the principle of juxtaposition and combination [26]. Most of these works are based on folk tunes, which are characterized by repeated themes. The change lies in the modification and enrichment of melodies by decorative sounds and changing sounds. The structural thinking is relatively simple and short. See Table 4 for the following examples.

“Lullaby,” a single part structure, consists of a segment and a’ segment. The length of the whole sentence is equal, and each paragraph is divided into (2 + 2). The a segment ends with a chord and belongs to an open structure segment.

The a’ segment is stably wrapped above the main chord, and D sentences are added on the basis of 4 sentence bodies to strengthen the sense of ending at the end of the music. The theme material is very simple. In the latter two sentences, the rhythm of attachment points is introduced to form a comparison with the previous materials. The ending uses the d-phrase, and the rear materials end in the way of voice part alternation. For Mongolian works, see Table 5 [27].

“Chulugen,” a single musical form, is composed of two single segments a and a’, and segment a is composed of two phrases a and B. Among them, the phrase a can be divided into two small music sections with 3 + 2 structure, and the phrase B has 2 + 2 music section structure. The material of the whole song is concentrated and runs through with “B feather mode,” which belongs to a single tonal music segment structure. See Table 6.

“Heyinghua” is a typical single phrase multi paragraph structure. Each phrase is 4 bars long with interlude in the middle. The introduction and excessive use of the same material. In the ‘B feather mode, other phrases are shown in the F angle mode. The form of introduction is simple and clear, which is the norm of folk music. The theme phrase consists of 2 + 2 stanzas. The latter stanza is like the answer sentence of the former stanza. It ends with a Shang tone and stays on the main chord in “D major,” giving people a sense of unfinished meaning and slight expectation. In the six theme presentations, the melody structure is relatively stable, and the main melody has been played by the first Matouqin. The changes of harmony and texture strengthen the audience’s memory of the theme [28].

In the traditional music of China and Mongolia, the horizontal line plays an absolutely dominant role. As a nation capable of singing and dancing, Mongolia has a large



TABLE 4: "Lullaby."

First order musical structure	Introduction	A	A <sup>1</sup>	Epilogue
Number of starting and ending sections	1~11	12~27	28~47	48~52
Secondary musical structure		a b c d	a <sup>1</sup> b <sup>1</sup> c <sup>1</sup> d <sup>1</sup> d <sup>2</sup>	
Number of sections	11	4 4 4 4	4 4 4 4 4	5
Modal tonality	<sup>b</sup> E major	C Feather		

TABLE 5: "Chulugen."

First order musical structure	Prelude	:A:	Connect	A <sup>1</sup>	Epilogue
Number of starting and ending sections	1~2	2~11	12	13~21	22~25
Secondary musical structure		a b		a <sup>1</sup> b <sup>1</sup>	
Number of sections	2	5 4	1	5 4	4
Modal tonality	<sup>b</sup> B feather				

TABLE 6: "Heyinghua."

Musical structure	Introduction	A	a <sup>1</sup>	Connect	a <sup>2</sup>	a <sup>3</sup>	Connect	a <sup>4</sup>	Transition	a <sup>5</sup>
Number of sections	4	4	4	2	4	4	2	4	3	6
Modal tonality	<sup>b</sup> B feather	Angle F							<sup>b</sup> B feather	

number of treasures of folk songs and folk music. In the continuous excavation and protection of predecessors, they are displayed in front of us in a variety of forms. In folk activities, the form of multi part music has already blossomed everywhere, but its form is generally relatively simple, focusing on the melody imitation of single part, and its creation follows the mode of linear thinking. In the long history of national music development, many singers and instrumental players have established simple vertical and horizontal sound combination experience through various ways, that is, what we call harmony today. Of course, some of the establishment of these acoustics are conscious, while others are unconscious. They often linger between regularity and irregularity. The composers of China and Mongolia rely on the nationality of Mongolian music in their creation, mostly around the pentatonic mode. After years of washing, people gradually no longer use simple hearing or experience to judge the quality of music, but pay more attention to the details of music itself. The appearance of the chamber music form of Matouqin ensemble is only in the past 30 years, and the creative groups are more complex, including composers, Matouqin players, conductors, and so on. In the frequent cultural exchanges between the two countries in recent years, the form of music is also constantly changing. More and more students majoring in composition in China choose to go to Mongolia for further study, and their style is also close to the creative style characteristics of Mongolia. In the study and application of harmony techniques, due to special historical reasons, China has interrupted cultural exchanges with Europe, America, and other countries, and implemented the policy of "leaning to one side" to the Soviet Union. Therefore, a large number of excellent works and

related books of the Soviet Union were widely spread in China. Sposobin harmony system had a far-reaching impact on the development of Chinese music. For Mongolia, which is adjacent to Russia, the concept of harmony is deeply influenced by it. It is more bold and open in creation and pays attention to the expression of diversified music ideas. Tone transfer is an indispensable and important technique in multi part music creation. Through the change of tonal color, we can express the music content and shape the music image. The power of harmony can be enhanced through tone transfer, which helps to promote the development of music and highlight the contrast and balance between various parts of music. Through tone transfer, with the help of tonal color change and functional function, harmony in music works can have rich color contrast and dynamic function. Composers in China and Mongolia both love the basic method of three degree superposition as chord composition in their creation, and are also committed to coordinating and expanding this harmonic method with the style of pentatonic mode. This has many similarities with China's pentatonic mode harmony theory. It can also be studied by using the national harmony theory proposed by Professor Fan Zuyin, such as several fixed harmony structure forms formed in the works, including harmony of three-dimensional structure, harmony of four and five-dimensional structure, and harmony of two-dimensional structure. In the chords that omit three tones, the omission of three tones leads to an empty and simple harmonic sound, which mainly appears in the form of texture or outer frame interval in the works. The chord structure with an additional 6 degrees has independent harmonic meaning. The use of a 6-degree interval attached to the triad is a common harmonic means in folk

music. In the way of tonality transformation, composers in China and Mongolia often prefer downward direction to tone. In Mongolia's works, tonality transformation is more frequent, but the overall tonal trend is still moving towards subordinates. Through the author's research, it is found that the technique of downward tone transfer has the foundation of music culture. In the long history of Mongolian music, the concept of tonality in folk songs has had a far-reaching impact on later composers, which is also the reason why most Mongolian music works prefer to turn down the tone.

## 5. Conclusion

This paper introduces the background, significance, and research status of music chord recognition. It can be understood that music chord recognition is an important research content in the field of music information retrieval. It involves the research of music theory, signal processing, machine learning, and artificial intelligence, and its application range is extremely wide, including music humming retrieval, audio detection and segmentation, music scoring system, and so on. At the same time, auditory model has been widely developed and applied in the field of music information retrieval in recent years, and achieved good results. Therefore, auditory model is applied to chord recognition in this paper. The experimental results show that this method has a certain development prospect. Firstly, the auditory image features of music chord segments are extracted, the one-dimensional music signal is converted into two-dimensional image features, and then the local features of auditory images, namely, SIFT feature vector, are extracted. Then, using the matching method of SPM, the local feature vectors of images are integrated into a feature vector to represent the features of complete auditory images, namely, the chord features of music. Secondly, the pattern recognition method based on sparse representation classifier (SRC) has achieved great success in image scene classification, target recognition, and face recognition. Then SRC has also been introduced into music genre, classical music classification, and music chord recognition, and has achieved good results. Therefore, this paper uses SRC method to recognize chords. Finally, the experiment is carried out under the optimal parameter setting. The experimental results show that the recognition effect of auditory image features and SRC recognition method is the best. The sounding principle of Matouqin is different from that of other musical instruments. Most of the instruments we see are bow string instruments. No matter how many strings there are, each string is composed of one string. For example, we are familiar with violin, Cello and guitar, as well as Chinese Erhu, and Tibetan string. But which musical instruments are different from the horse head Qin? Although the horse head Qin has only two strings, each string is composed of hundreds of horsetail wires. Due to the different length and tension of horsetail wire, the timbre of Matouqin will never be as "clean" as violin and erhu, but it is precisely this "unclean" that has become the unique timbre

of Matouqin different from other musical instruments and the root of its unique charm. In addition, musical instruments such as Sihu and Sanxian are also used by other ethnic groups. It seems that it is difficult to distinguish them from other ethnic instruments in some aspects, while the horse head Qin is unique and different from any musical instrument of other ethnic groups, so it has become the most representative musical instrument of Mongolia.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest with any financial organizations regarding the material reported in this article.

## Acknowledgments

This study was supported by Inner Mongolia Philosophy and Social Science Planning Project: Research on the Development Path of Morin Khuur Chamber Music Performance and Creation from the Perspective of Non-Genetic Inheritance, no. 2020NDC109.

## References

- [1] F. M. Shakiba and M. Zhou, "Novel analog implementation of a hyperbolic tangent neuron in artificial neural networks," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 10856–10867, 2021.
- [2] T. Zanotti, F. M. Puglisi, and P. Pavan, "Reconfigurable smart in-memory computing platform supporting logic and binarized neural networks for low-power edge devices," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 10, no. 4, pp. 478–487, 2020.
- [3] A. V. Liashok, Y. B. Popova, and P. Yuliya, "Using artificial neural networks to determine wear of composite friction material," *Science & Technique*, vol. 20, no. 4, pp. 345–351, 2021.
- [4] Z. Sabir, M. A. Z. Raja, M. Shoaib, and J. F. G. Aguilar, "Fmneics: fractional meyer neuro-evolution-based intelligent computing solver for doubly singular multi-fractional order lane-Emden system," *Computational and Applied Mathematics*, vol. 39, no. 4, p. 303, 2020.
- [5] S. Meftah, B. H. M. Tan, C. F. Mun, K. M. M. Aung, B. Veeravalli, and V. Chandrasekhar, "Doren: toward efficient deep convolutional neural networks with fully homomorphic encryption," *IEEE Transactions on Information Forensics and Security*, vol. 16, no. 99, pp. 3740–3752, 2021.
- [6] Z. Liu, X. Hu, T. Liu, X. Li, W. Wang, and F. M. Ghannouchi, "Attention-based deep neural network behavioral model for wideband wireless power amplifiers," *IEEE Microwave and Wireless Components Letters*, vol. 30, no. 1, pp. 82–85, 2020.
- [7] W. S. Monroe, F. M. Skidmore, D. G. Odaibo, and M. M. Tanik, "Hiho: accelerating artificial intelligence interpretability for medical imaging in IoT applications using hierarchical occlusion," *Neural Computing & Applications*, vol. 33, no. 11, pp. 6027–6038, 2021.

- [8] A. Ahf, B. Fm, and A. Ma, "Implementation of multilayer perceptron (mlp) and radial basis function (rbf) neural networks to predict solution gas-oil ratio of crude oil systems - sciencedirect," *Petroleum*, vol. 6, no. 1, pp. 80–91, 2020.
- [9] X. Wang, J. H. Park, S. Zhong, and H. Yang, "A switched operation approach to sampled-data control stabilization of fuzzy memristive neural networks with time-varying delay," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 3, pp. 891–900, 2020.
- [10] C. Carpenter, "Applications of artificial neural networks for seismic facies classification," *Journal of Petroleum Technology*, vol. 73, no. 02, pp. 68–69, 2021.
- [11] T. Batool, S. Abbas, Y. Alhwaiti et al., "Intelligent model of ecosystem for smart cities using artificial neural networks," *Intelligent Automation & Soft Computing*, vol. 29, no. 3, pp. 513–525, 2021.
- [12] D. Schneider, N. Korfhage, M. Mühling, P. Lüttig, and B. Freisleben, "Automatic transcription of organ tablature music notation with deep neural networks," *Transactions of the International Society for Music Information Retrieval*, vol. 4, no. 1, pp. 14–28, 2021.
- [13] S. A. Matveev, I. V. Oseledets, E. S. Ponomarev, and A. V. Chertkov, "Overview of visualization methods for artificial neural networks," *Computational Mathematics and Mathematical Physics*, vol. 61, no. 5, pp. 887–899, 2021.
- [14] M. Genaidy, "Using artificial neural networks models for predicting wheat yield productivity," *Arab Universities Journal of Agricultural Sciences*, vol. 28, no. 3, pp. 767–775, 2020.
- [15] N. Al Bitar and A. I. Gavrilov, "Using artificial neural networks to compensate for the error in an integrated navigation system," *Herald of the Bauman Moscow State Technical University Series Instrument Engineering*, vol. 2, no. 131, pp. 4–26, 2020.
- [16] N. Dar-Odeh, H. T. Fadel, O. Abu-Hammad, and O. A. Ghulam, "Are artificial neural networks useful for predicting overhanging dental restorations? a cross-sectional study," *World Journal of Dentistry*, vol. 11, no. 2, pp. 99–104, 2020.
- [17] J. S. Rahman, T. Gedeon, S. Caldwell, R. Jones, and Z. Jin, "Towards effective music therapy for mental health care using machine learning tools: human affective reasoning and music genres," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 11, no. 1, pp. 5–20, 2021.
- [18] G. N. Kamyshova, "Application of artificial neural networks for irrigation control," *The Agrarian Scientific Journal*, no. 4, pp. 84–88, 2021.
- [19] A. P. Singh, S. Dwivedi, and P. K. Jain, "A novel application of artificial neural network for recognition of target behind the wall," *Microwave and Optical Technology Letters*, vol. 62, no. 1, pp. 152–167, 2020.
- [20] T. Dedeken, "Research and development of an artificial neural network for spectral data," *ACC Journal*, vol. 26, no. 1, pp. 19–30, 2020.
- [21] J. D. Olden and D. A. Jackson, "Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks," *Transactions of the American Fisheries Society*, vol. 130, no. 5, pp. 878–897, 2001.
- [22] B. R. Szkuta, L. A. Sanabria, and T. Dillon, "Electricity price short-term forecasting using artificial neural networks," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 851–857, 1999.
- [23] X. Yao, "A review of evolutionary artificial neural networks," *International Journal of Intelligent Systems*, vol. 8, no. 4, pp. 539–567, 1993.
- [24] H. White, "Learning in artificial neural networks: a statistical perspective," *Neural Computation*, vol. 1, no. 4, pp. 425–464, 1989.
- [25] M. Akkaya and A. Haydar, "Analysis of job satisfaction of academicians through the use of artificial neural networks," *Academic Research International*, vol. 4, no. 1, p. 721, 2013.
- [26] R. Andrews, J. Diederich, and A. B. Tickle, "Survey and critique of techniques for extracting rules from trained artificial neural networks," *Knowledge-Based Systems*, vol. 8, no. 6, pp. 373–389, 1995.
- [27] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, "Artificial neural networks in hydrology," *Journal of Hydrologic Engineering*, vol. 5, no. 2, pp. 124–137, 2000.
- [28] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996.