ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Analysis of the Spreading Path of Chinese National Musical Instruments on the World Music Stage

Yaju Li[1,*]

[1] School of Education, Pingdingshan University, Pingdingshan, Henan, 467000, China

SUMMARY: *The present paper uses a data-driven research paradigm to conduct a systematic examination of the distribution channels of Chinese traditional musical instruments around the world music scene. Initially, it will be done through acoustic analysis of audio samples by extracting and refining audio features according to the MPEG-7 standard (spectral centroid, spread spectrum, and their first- and second-order derivatives) using the K-Nearest Neighbors (KNN) classification algorithm to recognize instruments automatically. On top of this base, a retrieval system has been built to reach efficient retrieval of target instrument sample out of global music databases. Another concept of agent-based modeling was presented and the classes of agents were defined: propagators, receivers, and immune agents with their interaction rules and the Runge-Kutta method was used to perform numerical simulation. The results indicate that adding dynamic eigenvalues is much useful in improving classification performance where the mean F1 score increases by 86.29% to 96.42%. Moreover, experimental harmonic structure analysis determined the best harmonic order to be 7, which increased the recognition rate of all four categories of instruments to over 93.Finally, to simulate the propagation mechanism, the equilibrium stability of musical information transmission was revealed: when the basic reproduction number* $R_0 = 3.5$*, the system stabilizes after approximately 8 days. The study further demonstrated the significant impact of knowledge agents' individual characteristics across different network layers (In, Middle, Out) on knowledge diffusion rates and system evolutionary dynamics. This research provides novel insights and tools for promoting Chinese traditional musical instruments globally.*

KEYWORDS: *traditional musical instruments; MPEG-7; music information retrieval; agent-based modeling; knowledge dissemination; Runge-Kutta algorithm*

1 Introduction

In today's increasingly accelerated globalization, cultural confidence has become an important part of a country's soft power [1, 2]. Chinese folk musical instruments have a long history, and with the evolution of society and cultural exchanges, musical instruments have gradually become rich and diverse, and become an indispensable part of Chinese culture [3]. There are many types of Chinese folk musical instruments, which can be roughly categorized into wind instruments, plucked instruments, percussion instruments and other major categories according to the different ways of playing [4, 5]. Wind instruments mainly include flute, xiao, sheng, suona, etc. These instruments are pronounced by blowing vibration, which is characterized by melodious tone and rich expressiveness. Among them, the flute is popular for its clear and pleasant tone and rich playing skills, while the suona is famous for its high and exciting tone

*bella_star88@163.com
https://doi.org/10.65102/is2026396
Published: 30 April, 2026

and strong infectious force. Plucked instruments mainly include pipa, guzheng, yangqin, etc. These instruments are played by plucking the strings with fingers or plectrums to pronounce sounds, and are characterized by bright tones and strong sense of rhythm. Among them, the pipa is one of the representatives of ancient Chinese music for its rich playing skills and unique sound, while the guzheng is popular for its long history and beautiful sound. Percussion instruments mainly include drums, cymbals and cymbals, etc. These instruments are characterized by a strong sense of rhythm and shocking sound by striking the sound. In ancient wars, percussion instruments were often used to boost morale; in modern music, they have become an important tool for setting the mood.

Chinese folk musical instruments are closely related to traditional culture and folklore, and they are not only a tool for music performance, but also a carrier of cultural inheritance [6]. Many ancient literary works and historical stories are spread and interpreted by musical instruments, which makes the cultural connotation of musical instruments richer and deeper. At the same time, the position of musical instruments in traditional culture is also very important. In ancient rituals, celebrations and other activities, musical instruments are one of the indispensable elements, which can not only set the atmosphere and express emotions, but also convey cultural information and values [7, 8].

In addition, Chinese ethnic musical instruments are also closely related to regional culture and ethnic customs. The styles and playing methods of musical instruments vary from region to region and from ethnic group to ethnic group, and these differences not only reflect the diversity of cultures, but also enrich the musical treasury of the Chinese nation [9, 10]. However, with the changes of the times, especially the impact of popular music, the inheritance and development of traditional musical instruments are facing unprecedented challenges, which makes the younger generation more inclined to accept modern forms of music, while their interest in and knowledge of traditional musical instruments are gradually decreasing [11, 12].

At present, the most effective path for the dissemination of Chinese national musical instruments is to utilize the digital intelligence technology, through the Internet in the dissemination, to enhance the popularity and influence of Chinese national musical instruments, and then promote their inheritance and dissemination [13]. Literature [14] shows Chinese folk musical instruments through rich and vivid audiovisual language, reveals the development history, humanistic feelings and production process of traditional Chinese musical instruments, and explores the role of the movie “Chinese Musical Instruments” in the dissemination and protection of Chinese folk musical instruments. Literature [15] points out that the use of network platforms, multimedia technology and other tools for the dissemination of Chinese folk music and musical instruments has a certain degree of promotion, through multi-channel, three-dimensional dissemination strategy to achieve globalization and sharing of Chinese national cultural connotations. Literature [16] developed an AR application based on education and tourism in order to proliferate the recognition of Chinese national musical instruments in the market, and the program greatly attracted users' interest in Chinese national musical instruments and strengthened their understanding of Chinese national culture. Literature [17] proposed an interactive system for chimes based on digital augmented projection and somatosensory technology, which comprehensively showed the multi-dimensional scenes of chimes production and performance, provided users with a personalized and diversified interactive platform for chimes, and helped the chimes spread on the world music stage. Literature [18] designed a multi-sensory digital musical instrument (OrigamiSOUND) based on the dynamics and interactivity of digital technology, and combined with the results of usability scales, participation scales and behavioral analyses, they all proved the usability of OrigamiSOUND, as well as the fact that OrigamiSOUND improves the user's knowledge of the culture of Chinese national musical instruments. Literature [19] comprehensively analyzed the cultural characteristics of traditional Chinese ethnic musical instruments in Shandong Province, China, through the field survey method, combining quantitative and qualitative research methods, based on which effective paths for the promotion of Chinese ethnic musical instruments, such as the establishment of music education centers, digital learning platforms, collaborative performances, and music festivals, were proposed. Literature [20] utilized multimedia technology to build a database of Chinese musical instruments, which provides textual descriptions and audio-visual displays of more than 2,000 musical instruments, effectively enhancing the user's awareness of Chinese musical instruments, and has great potential for application in spreading the culture of Chinese musical instruments. Literature [21], in order to promote the traditional musical instrument culture in Xinjiang, considers that traditional musical instruments are digitally processed and digital museums are constructed, through which all kinds of cultural information of traditional musical instruments in Xinjiang are transmitted to people in order to disseminate the culture of ethnic musical instruments in Xinjiang.

Digital intellectualization technology can record the sound of ethnic musical instruments in high-fidelity audio format, and with the help of professional audio and video playback equipment, it can highly restore the real scene of musical instrument playing and promote the dissemination of Chinese ethnic musical instruments [22]. With the help of the Internet and digital media platforms, the audio and video materials of ethnic musical instruments can be rapidly disseminated to various places, and they can be easily accessed through the Internet, whether by professional music researchers or by the general public interested in ethnic music [23, 24]. It breaks the limitations of traditional dissemination methods in terms of geography and time, and enlarges the influence and audience range of ethnic musical instruments.

The article starts from the microscopic recognition of audio features of musical instruments and gradually advances to the macroscopic simulation of music communication dynamics, forming a data-driven, layer-by-layer methodological system. Firstly, we analyze the acoustic features of ethnic musical instruments to accurately capture the physical nature of instrumental timbre. On this basis, a classification method based on timbre database (K-nearest neighbor algorithm and improved MPEG-7 feature extraction) is introduced to realize automated identification and categorization of different ethnic musical instruments, which provides support for the construction of a structured musical instrument timbre database. Then, in order to effectively retrieve the music samples containing the target ethnic instruments from the complicated global music database, we continue to construct a retrieval system based on the music analysis model, which elaborates the whole process from the construction of the retrieval engine to the construction of the index based on the musical note sequences, which ensures the accuracy and efficiency of the acquisition of research samples. Finally, rising to the macroscopic simulation of the propagation mechanism, the intelligent body-based modeling method is introduced, and a computable and simulatable music propagation model is constructed by defining the intelligent body classes, environments, and their interaction rules, which can be used to dynamically decipher the acceptance, integration, and diffusion paths of the Chinese ethnic musical instruments in the international arena.

2 Technological analysis model for the dissemination of Chinese folk musical instruments

2.1 Acoustic Characterization of Audio

Firstly, the acoustic features of Chinese ethnic musical instrument audio are extracted and quantized.

According to the latest research field, MPEG has released a set of feature descriptors for describing digital audio, both for the time and frequency domains, and the Short Time Fourier Transform (STFT) with a Hemming window generates an instantaneous value at each frame of these feature attributes.

(1) Spectral center of mass

The spectral center of gravity of audio is used to describe the center of gravity of the spectral power spectrum, which indicates the predominant frequency range as a factor of the power spectrum. The coefficients below 62.5 Hz have been merged for fast computation.

(2) Spread Spectrum

The spread spectrum is the mean square value of the logarithmic power spectrum of the center of gravity for each frame, which is used to describe the shape of the power spectrum. The standard deviation of the mean of all frames is used to describe the spread spectrum of the audio signal.

2.2 Classification methods based on timbre databases

In order to systematize the above audio acoustic features and achieve automatic identification of musical instruments, this section introduces a classification method based on a timbre database to provide structured support for subsequent propagation studies.

2.2.1 K-nearest domain taxonomy

Nearest Neighbor Classification is a learning method based on analogy, i.e., learning by comparing a given test instance with training instances that are similar to it. The training instances are described by n attributes, and each instance represents a point in the n-dimensional space, such that all training instances are put into the n-dimensional pattern space. When given an unknown instance, the k nearest domain taxonomy searches this pattern space to find the k training instances of the nearest unknown instance. These k training instances are the k “nearest neighbors” of the unknown instance.

2.2.2 Improved feature value extraction method based on MPEG-7

Currently, extracting spectral features of signals to describe timbre information is used as the main method to identify different musical instruments. In order to describe power spectral features, MEPG-7 proposes a number of useful and relevant timbral features, such as spectral plosives and spreading. In addition, the transient temporal characteristics of music provide some relevant timbral features which complement the spectral features to fully represent the timbral quality of the sound.

In this paper, two continuous spectral features of the spectral center of mass and spread spectrum are calculated directly to obtain new eigenvalues. These new eigenvalues are calculated as

$$C_i{}' - \left(C_{i+1} - C_i\right) / C_i \tag{1}$$

$$S_i{}' - \left(S_{i+1} - S_i\right) / S_i \tag{2}$$

C_{i+1}, C_i and S_{i+1}, S_i are two consecutive frames of the spectral center-of-mass and the spreading frequency, respectively, and C_i' and $S_i{}'$ are the rates of change of them, respectively. Following the same approach, the rate of change of C_i' and S'_i can be calculated, which can usually be thought of as the second-order derivatives with respect to the spectral center of mass and the spread. The formula is

$$C_i'' = \left(C_{i+1}' - C_i'\right) / C_i' \tag{3}$$

$$S_i'' = \left(S_{i+1}' - S_i'\right) / S_i' \tag{4}$$

Obviously, C_i' and S_i' inscribe the temporal variations of the two neighboring spectra, and C_i'' and S_i'' further inscribe the temporal variations of every three neighboring spectra. In the later experiments, the new eigenvalues are added to the experimental data, and the classification algorithm is used to specifically analyze whether the eigenvalues based on the new temporal characteristics can yield better results in terms of two criteria: recall and accuracy.

2.3 Music retrieval based on music analysis modeling

The establishment of classification methods has enabled the effective archiving of ethnic musical instrument timbres, but in order to quickly locate related works in global music databases and analyze their dissemination trajectories, an efficient music retrieval system needs to be further constructed. Therefore, this section explores the retrieval techniques based on music analysis models to realize the accurate extraction of Chinese folk instrument music samples from massive data.

2.3.1 Music Search Engine Construction

The workflow of the basic retrieval engine is shown in Fig. 1, where firstly, text data, audio data, and sheet music data (picture format input) are processed to obtain a query index, according to which the query index is looked up in the database, and after that the retrieval results are sorted, and the retrieval results that satisfy the conditions are output. The retrieval engine requires a high overhead, and for the processing of audio and sheet music, two types of processing are involved, i.e., the combination of audio feature extraction and the recognition and processing of sheet music images. In addition, the accuracy of the recognition and processing of the sheet music images is not high and depends on the image quality and the recognition algorithm.

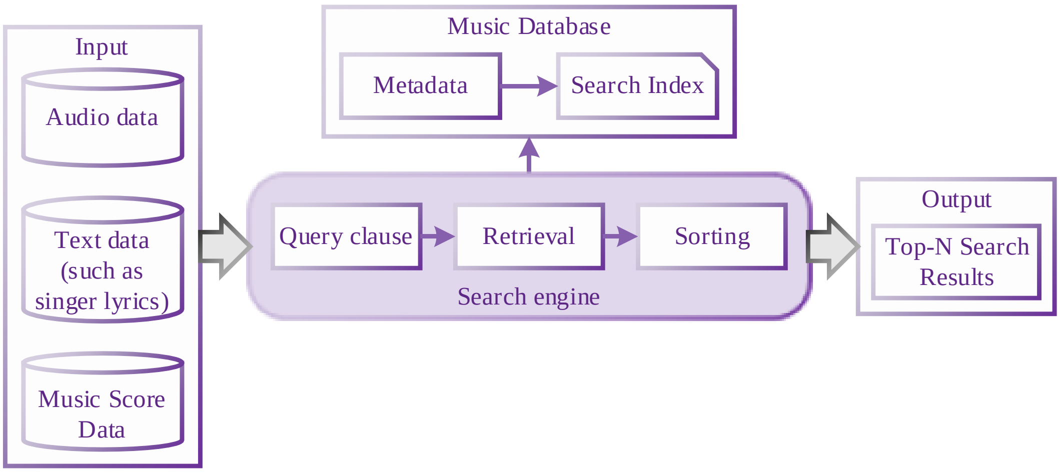

Figure 1: The flow chart of music retrieva

On this basis, this paper combines the music analysis model and similarity calculation method to propose an improved retrieval engine to process multiple types of input data to obtain retrieval results. The specific processing flow is shown in Figure 2.

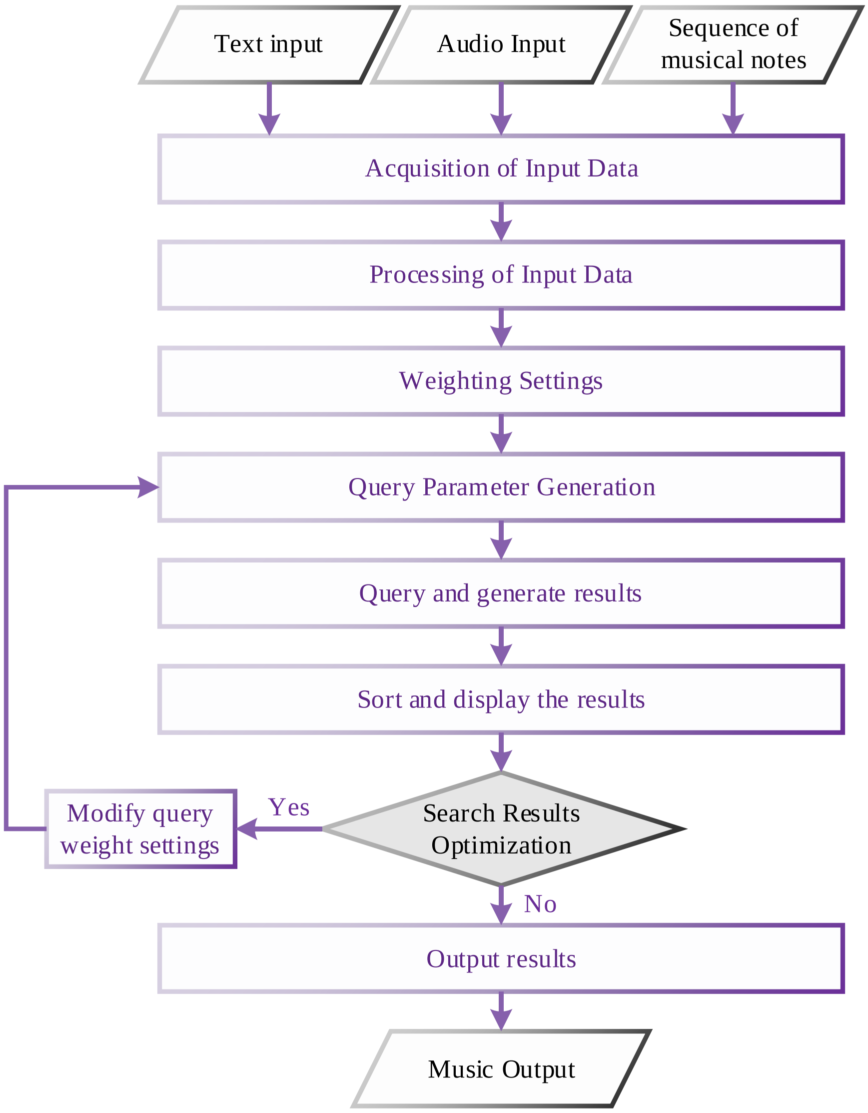

Figure 2: The flow chart of music search engine processin

The main process of the music retrieval engine is: firstly, the input is obtained, which is divided into three types of inputs: text, audio, and note sequences. Secondly, the acquired input data are processed, including text parsing, audio extraction of note sequences, normalization of note sequences, etc. After that, for the extracted data, the weighting is set according to the input parameter types to obtain the query parameters. Then, the data in the database is queried and the results are generated based on the query parameters. Finally, the results are sorted and the results are displayed. For the queried results, the user is allowed to choose whether to optimize or not, if no optimization is needed to directly output the music according to the corresponding format, if the user chooses to continue to optimize, the query weight setting can be modified, the query parameters can be regenerated and queried again.

The work of the retrieval engine relies on the construction of the music database and the pre-processing of the data in the database, using the methods mentioned in the previous sections of the paper to process the three types of data and store the vector model obtained from the processing. For the original audio signals, lyrics data and sheet music information, the music analysis model is first processed and built for storage, while the correspondence between the original data and the analysis model is retained and associated through indexing. In music retrieval, the music analysis model is first retrieved through the index, and then the original music file of the retrieved information is obtained according to the association index. The final data returned is the original format data file.

In summary, this subsection proposes a retrieval engine based on music analysis model to preprocess the music library data, which can get higher efficiency in the actual retrieval phase. Meanwhile, the query parameters are weighted and set during retrieval to qualify for different types of retrieval, which improves the accuracy of retrieval.

2.3.2 Index construction process based on note sequences

According to different types of music data can be constructed a variety of types of indexes, this subsection focuses on the construction process of indexes based on note sequences, and accordingly realizes the retrieval method based on note sequences, which extracts the note sequences from the input audio, and then retrieves them according to the note sequences, and also supports retrieval according to the note sequences input by the user.

First, the note sequence input of the note sequence-based retrieval method is discussed. According to the note categories, each note category has seven fetches, which are do, re, mi, fa, so, la, xi. Each fetch has two parameters, which represent pitch and duration, i.e., a note can be represented as note(frequency, duration). Among them, frequency represents the relative frequency, the base frequency is set to 264Hz, which represents the center octave, at this time, frequency is set to 5, duration is set by default to use 48 to represent a beat, can be converted to get the value, the requirement of the data format is an integer.

$$seq = so(5,48), re(5,48), fa(5,48), so(5,48) \tag{5}$$

For example, the input sequence as shown in Equation (5) represents a musical fragment whose tones are all in the center octave, each tone occupying one beat, and four beats in length. The note sequence can express the melodic information of the music more completely.

For retrieval, it is necessary to set the maximum or minimum length threshold of the input sequence; if the note sequence is too short, the number of music pieces that meet the requirements is too large, and the subsequent sorting and screening is more difficult; if the sequence is too long, the time complexity required for matching the note sequences is high, and it cannot satisfy the retrieval speed requirements.

In addition, for the data in the music database, the generation of note sequences is carried out and stored in advance, preserving the correspondence from note sequences to musical scores. A backward index of notes to note sequences is established, and an event table of indexed notes is also established, pointing to the relative position of the notes in the musical score represented by the note sequences, and in each position (number, index), number denotes the musical score number, and index denotes the position of the note in the musical score.

2.4 Interpretation of the Intelligentsia-based model of music communication

Understanding how instrumental music spreads and evolves in different cultural environments requires dynamic simulation tools. In this section, we introduce an intelligent body-based music dissemination model to simulate the dissemination and interaction mechanism of Chinese ethnic musical instruments on the world stage through computational methods.

The process of music information dissemination is similar to viral dissemination, then the music information disseminator is practicing music information dissemination, the music information receiver is able to receive the music information, and eventually becomes a new music disseminator. However, music receivers may also choose not to receive music communication and become immune. Therefore, in this paper, the music propagation intelligences in the music propagation intelligences are set as C, the receivers are set as D, and the immunizers are set as E when music propagation is practiced. Set the average ratio of music propagation intelligences that receive music information per unit of time as α, the ratio of music propagation intelligences that reject music information per unit of time as β, and the average ratio of music propagation intelligences that receive music information per unit of time as j, then:

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -\alpha jC \tag{6}$$

$$\frac{\mathrm{d}j}{\mathrm{d}t} = \alpha jC - \beta_j \tag{7}$$

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = \beta_j \tag{8}$$

where $\mathrm{d}/\mathrm{d}t$ denotes the derivation.

2.4.1 Model assumptions

(1) In a music dissemination environment, there exists only one source of music information dissemination and this source is a member of the music dissemination environment. This member is practicing music dissemination to nearby music dissemination intelligences when practicing music information dissemination.

(2) The music information dissemination time is subdivided so that at most one music information dissemination occurs between music disseminating intelligences in one unit of dissemination time.

(3) The initial number of nodes, the degree of willingness to propagate, the information value parameter, the moderating coefficient, and the situation of the nodes in the music propagation environment are not unknown.

2.4.2 Classes of music communication intelligences

According to the interaction process of music communication intelligences in the music communication environment, this paper sets the attributes of the music communication intelligences class, which contains the degree of willingness to communicate music, the chance of communication, the chance of re-communication, the number of times of communication, and its own rank.

Propagation Willingness Degree: describes the propagation willingness of the music propagation intelligence j for music information at the moment of e. After practicing music information dissemination, the dissemination willingness function is:

$$\varpi_j(e) = \varpi_j(e-1) + \phi \cdot \left(d^{e-eC} - 1\right) \cdot \left[\varpi_j(e-1) \cdot \varpi_j(e-1)\right] \tag{9}$$

In the equation: The willingness to transmit music information among recipients is denoted as $\varpi_j(e)$; ϕ represents the music information value parameter, where $\phi \in [0,1]$; The music information transmission willingness decay function is set as $\left(d^{e-eC} - 1\right)$, where eC denotes the number of times the music information has been transmitted.

Propagation probability: Describes the probability that a music propagation agent achieves music information propagation at time e within the music propagation environment. To reduce model complexity, $q_1(e)$ is used to describe the propagation probability function. When executing music information propagation, the probability function for a propagator's successful transmission is:

$$q_1(e) = 1 - \left(1 - \varpi_j\right) \tag{10}$$

In the music propagation environment, when music propagation agents execute music propagation, the probability of successful propagation is primarily affected by differences in propagation willingness between the propagator and the receiver. Due to variations in propagation willingness, the propagation probability function also exhibits differences.

Re-transmission Probability: Describes the probability that an immune agent within the music transmission agent network will re-execute music information transmission at time e. Thus, when executing music information transmission, the probability function for an immune agent to re-execute transmission is:

$$q_2(e) = 1 - \left(1 - \varpi_\beta\right) \tag{11}$$

In the equation, ϖ_β represents the transmission willingness function of immune agents.

If music transmission agent j in the music transmission environment belongs to the immune group, then the probability of agent j transmitting again within the environment is related to its own transmission willingness. If $q_2(e)$ exceeds the preset retransmission threshold V', this music transmission agent may retransmit music information; otherwise, it will not transmit music information.

Number of Transmissions: Describes the number of times a music transmission agent interacts with other agents to transmit music information at time e. This is denoted as $M(e)$ in this paper.

Self-Rank: Describes the rank of a music transmission agent within the music transmission environment, denoted as $H(e)$. $H(e) = \{1, 2, 3, 4, 5\}$. When a music propagation agent interacts within the environment, the value of $H(e)$ is decisively influenced by interactions with other agents. If $H(e) = 5$, it indicates the highest rank within the environment.

2.4.3 Music Dissemination Environment Category

This paper defines the attribute of the music transmission environment as participation rate.

The participation rate describes whether music transmission exists in the music transmission environment at time e, denoted as $D(e)$, where the value of $D(e)$ is either 0 or 1. If $D(e) = 1$, it indicates that music information transmission exists in the music transmission environment at time e. If $D(e) = 0$, it indicates that no music information transmission exists in the music transmission environment at time e.

2.4.4 Communication Interaction Rules Between Agent Classes

During the propagation interaction between music propagation intelligences, the propagator j in a music propagation intelligence practises music propagation to a nearby music propagation intelligence i, if the degree of willingness to propagate of j is greater than the degree of willingness to propagate of i and the chance of successful propagation of the information q_1 is not less than the set propagation threshold V, then music propagation intelligence i belongs to the new propagator; if the music information is rejected, then the music propagating intelligences i belong to the immunizer. The propagation chance q_1 between music propagating intelligences is interfered by the degree of propagation willingness of the two parties who propagate music information.

If the music propagation intelligent body i receives the music information propagation, its own propagation willingness degree is implemented according to the propagation willingness degree function, then the propagation number $M_i(e) = M_i(e-1)+1$.

When the music propagation intelligence interacts with the music propagation environment, the music propagation environment generally has a greater influence on the propagation of music information. If j's own rank is less than a specific threshold value $V(e)$ when the music propagation intelligence j is practicing music information propagation to the nearby music propagation intelligence i, music information exists in the music propagation environment where the music propagation intelligence is located, and vice versa, music information does not exist in the music propagation environment.

If music information exists in the music propagation environment, the music propagation environment will propagate the music information to multiple music propagation intelligences, if $q(e)$ is not less than the set propagation threshold V, the music information receiver becomes a music information propagator; if it refuses to receive the music information, the music propagation intelligence belongs to the immunizer. The chance of a music propagating intelligence to receive music information from the music propagating environment is mainly interfered by the degree of willingness of the music propagating intelligence to propagate.

3 Experimental validation of timbre characterization and recognition model for ethnic musical instruments

The system constructs technical analysis models from audio feature extraction, instrument classification to music retrieval and propagation simulation. In order to verify the effectiveness of these models, especially the improved feature extraction methods and classification algorithms based on MPEG-7, a series of experiments will be carried out in this chapter to empirically analyze the performance of automatic identification of ethnic musical instrument timbre.

3.1 Experiments on Musical Instrument Classification Based on MPEG-7 Improved Eigenvalues

3.1.1 Experimental database construction

The audio files used in this paper consist of audio data performed by twelve types of ethnic instruments provided by a certain orchestra. These twelve ethnic instruments can be subdivided into four categories:

Bowed instruments: Erhu, Zhonghu, Gaohu, Morin khuur;
Plucked instruments: Pipa, Guzheng, Ruan;
Percussion instruments: Yangqin, Gong, Drum;
Wind instruments: bamboo flute, xiao.

Feature values extracted from these instruments' audio files based on the MPEG-7 standard serve as instance samples. Each ethnic instrument includes 500 audio files, divided into training and testing sets at an 8:2 ratio.

For data selection, audio files were chosen in WAV format with a sampling rate of 45.2 kHz and 16-bit precision. Short-term energy-based dual-threshold endpoint detection method was used to remove silent segments.

In the chosen monophonic timbre dataset, all monophonic samples in it were processed. Longer than 3 second segments were cut, whereas shorter than 3 second segments were stretched by looping. Tables 1 shows the final monophonic database of ethnic instruments used in the experiment.

Table 1: Ethnic Instrument Monophonic Database

Category	Instruments	Number of single tones
Stringed instruments	Erhu	117
	Zhonghu	102
	Gaohu	136
	Maotuqin	86
Stringed instruments	Pipa	140
	Guqin	151
	Nguyen	73
Percussion instruments	Yangqin	94
	Lantern	112
	Drum	128
Blowing instruments	Flute	141
	Xiao	129

The information indicates that this database has 12 kinds of instruments and 1,489 monophonic sounds which are divided in four broad groups such as bowed, plucked, struck, and wind instruments. Regarding sample size distribution, the number of single note samples is not equal among different instruments. As an example, the zither (151 samples), bamboo flute (141 samples) and pipa (140 samples) have the largest sample sizes, whereas the ruan (73 samples) and horsehead fiddle (86 samples) have relatively small sample sizes. Such distribution indicates both the differing levels of difficulty with which instruments may produce stable and independent single notes in performance, and the logistical limitations to the data collecting process. The database offers a strong base upon further feature extraction and classifier training with its categories and quantities set up to fully evaluate model recognition behaviour with real-life unbalanced data.

3.1.2 Confusion Matrix

Figures 3 and 4 illustrate the classification confusion matrices of the audio of each instrument in Experiment 1 and Experiment 2 respectively, both before and after the addition of new features.

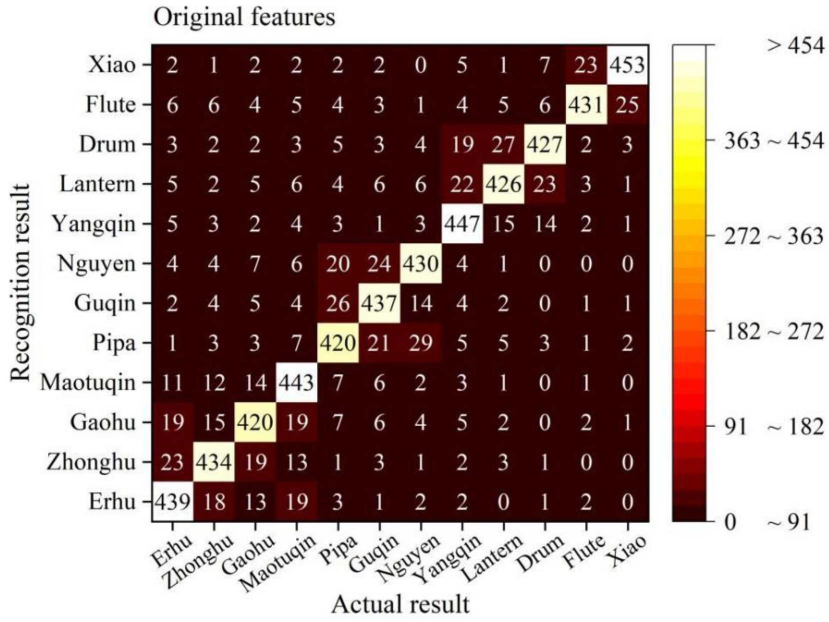

Figure 3: Confusion matrix of traditional eigenvalues

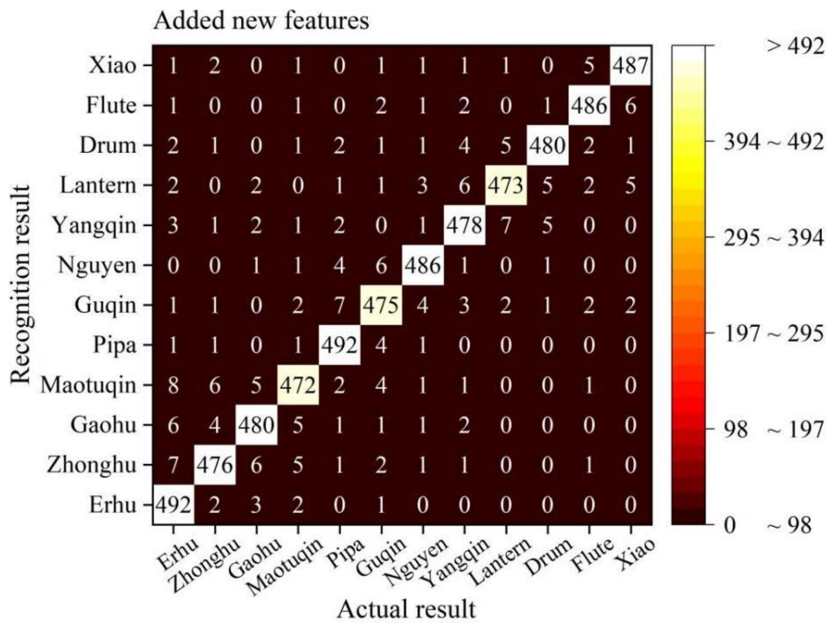

Figure 4: Confusion matrix with added new feature values

The comparison between the confusion matrices before and after introducing new feature values shows that the addition of these new values will greatly enhance the precision of instrument classification.

The number of correct classification of instruments based on categories was usually small when using traditional features, and there were many misclassifications. E.g., the erhu was correctly classified 439 times but erroneously classified as zhonghu 18 times, gaohu 13 times, and horsehead fiddle 19 times; zhonghu was correctly classified 434 times but erroneously classified as erhu 23 times and gaohu 19 times. Percussion instruments such as gongs and drums had significant classification errors. The number of gongs misclassified as drums was 23 and the number of drums misclassified as gongs was 27 which shows that traditional features are not able to classify instruments with similar timbres well. In general, the confusion matrix at the traditional features revealed a broad distribution of errors, particularly large misclassification rates between the same-category instruments.

The inclusion of new feature values led to a considerable increase in the number of correctly classified instruments in all categories. In the case of erhus, the correct number of classifications increased to 492 (up from 439) and the number of misclassified samples was also greatly reduced, e.g., misclassification as zhonghu decreased by 18 (to only 2), and misclassification as morin khuur decreased by 19 (to only 2). Likewise, correct classifications of zhonghu went up to 476 (434) and gaohu to 480 (420) and morin khuur to 472 (443). Correct classifications of plucked instruments such as pipa grew by 420 to 492, guzheng grew by 437 to 475, and ruan grew by 430 to 486. Clasification results on percussion instruments such as yangqin, gong, and drum were also enhanced significantly. Gong misclassification as drum has been reduced by 23 to 5 cases and drum misclassification as gong has been reduced by 27 to 5 cases. Accuracy in classification of bamboo flute and xiao also increased, with the bamboo flute having a count of correct classifications of 431 to 486 and the xiao having an increase of 453 to 487.

The new feature values represented temporal changes in the audio spectrum, including rate of change between consecutive frames and higher-order derivatives, which increased the dynamic nature of instrument timbre differentiation. The experimental findings prove that these temporal attributes are very efficient in minimizing confusion between similar instruments and enhance classification performance. In general, the introduction of additional features tended to increase the diagonal elements of the confusion matrix and greatly decrease the off-diagonal errors, thus demonstrating the effectiveness of the MPEG-7-enhanced feature-based approach to automatic recognition of traditional Chinese musical instruments.

3.1.3 Comparing Classifier Performance

The Precision, Recall and their harmonic mean F1 score are used in this paper to evaluate the performance of the classifier based on the classification performance of 12 musical instruments pre and post the addition of new features. The particular performance measures are given in Table 2. To further show how the classification has improved upon adding new features, line charts of the F1 scores of both experiments are depicted in Figure 5. The shaded areas denote the F1 score improvement of every instrument with and without new features added.

Table 2: Classification performance of 12 types of musical instrument audio

	Original features			Added new features		
	Precision	Recall	F1	Precision	Recall	F1
Erhu	84.42	87.48	85.95	94.89	98.59	96.74
Zhonghu	86.11	86.28	86.20	96.36	95.28	95.82
Gaohu	84.68	84.23	84.46	96.19	96.73	96.46
Maotuqin	83.43	88.61	86.02	95.93	95.21	95.57
Pipa	83.67	84.53	84.10	96.09	98.46	97.28
Guqin	85.19	87.43	86.31	95.38	95.08	95.23
Nguyen	86.69	86.23	86.46	97.01	97.85	97.43
Yangqin	85.63	89.04	87.34	95.79	95.82	95.81
Lantern	87.33	85.12	86.23	96.93	95.17	96.05
Drum	88.59	85.43	87.01	97.36	96.18	96.77
Flute	92.09	86.52	89.31	97.39	97.67	97.53
Xiao	93.02	90.67	91.85	97.21	98.19	97.70

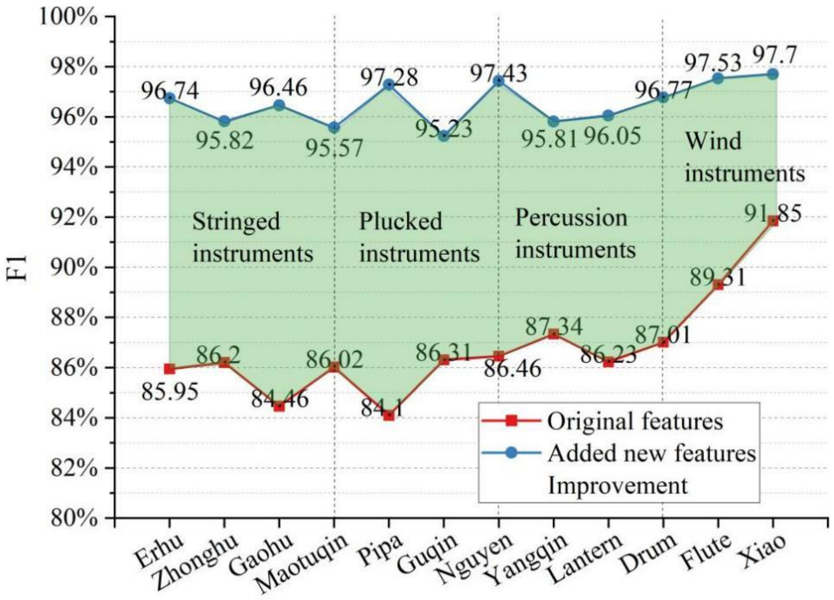

Figure 5: F1 values before and after adding new features to each instrument

Overall, the addition of new features resulted in comprehensive and significant improvements across all three performance metrics for all instruments.The mean F1 value with conventional features was around 86.29, and it rose up to 96.42 when the new features were included, an increment of greater than 10 percentage points. It is evident that the newly derived temporal variation features are crucial in determining the dynamism of instrumental timbre and increasing the discriminatory abilities of the model.

3.2 Harmonic Order Optimization Experiment

The stated experiments support the importance of adding time dynamic properties to the improvement of precision of music instrument classification. Nevertheless, the depiction of timbre attributes is not only dependent on dynamic data but also depends on harmonic structure as one of the main elements. In this section, the methods of feature extraction using harmonic structure are discussed in more detail and the best harmonic order is identified by conducting experiments to optimally optimize the timbre recognition model.

3.2.1 Experimental Data Preprocessing

The classification experiments are still being conducted with the MPEG-7 standard audio files that contain 12 Chinese ethnic instruments in the mentioned dataset. The data that should be predicted is to predict bowed string music with the erhu, zhonghu, gaohu, and morin khuur. Waveform diagrams of them are presented in Figure 6.

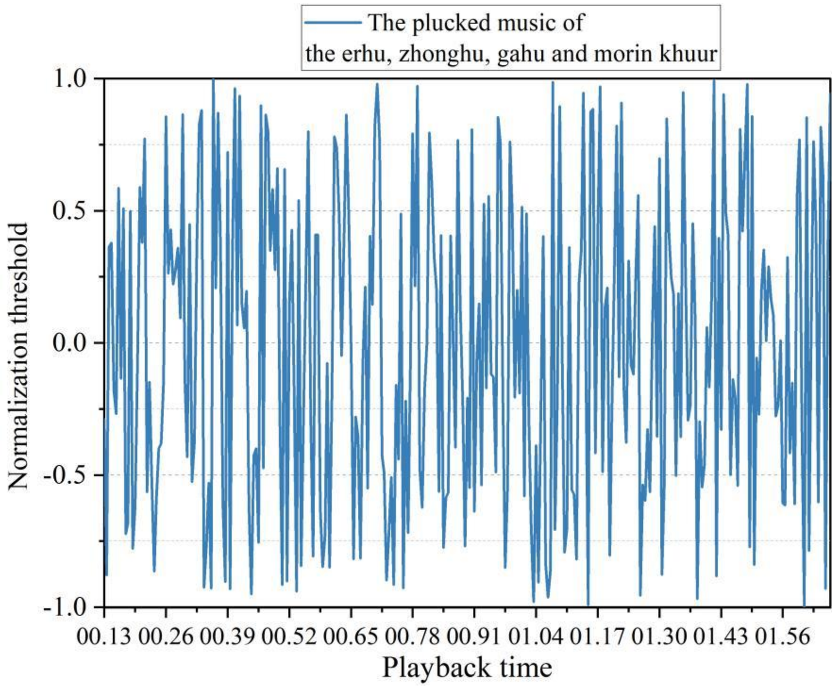

Figure 6: The chord music waveforms of the erhu, zhonghu, gaohu and maotuqin

The waveform diagram is a visual representation of the acoustic superposition effect during ensemble playing. Stringed instruments generally have continuous periodic oscillations in their waveforms whereas ensemble playing consists of combining the fundamentals of different instruments with their respective harmonics to produce a more complex composite waveform. The diagram shows that the amplitude envelope varies, indicating that the dynamics are changing because of the typical vibrato technique of the stringed instruments which has a distinctive dynamic change.

3.2.2 Fundamental Frequency Detection Based on the Autocorrelation Function

The period of the fundamental frequency of an erhu audio signal is determined based on the properties of the short-term autocorrelation function. When comparing the likeness of the original signal to the delayed signal, the maximum similarity occurs when the delay between them is equal to the fundamental frequency period. Or else, the distance between the two signals with the greatest similarity is also equivalent to one fundamental frequency period.

Figure 7 shows a single-note signal from an erhu and a frame of its musical signal.

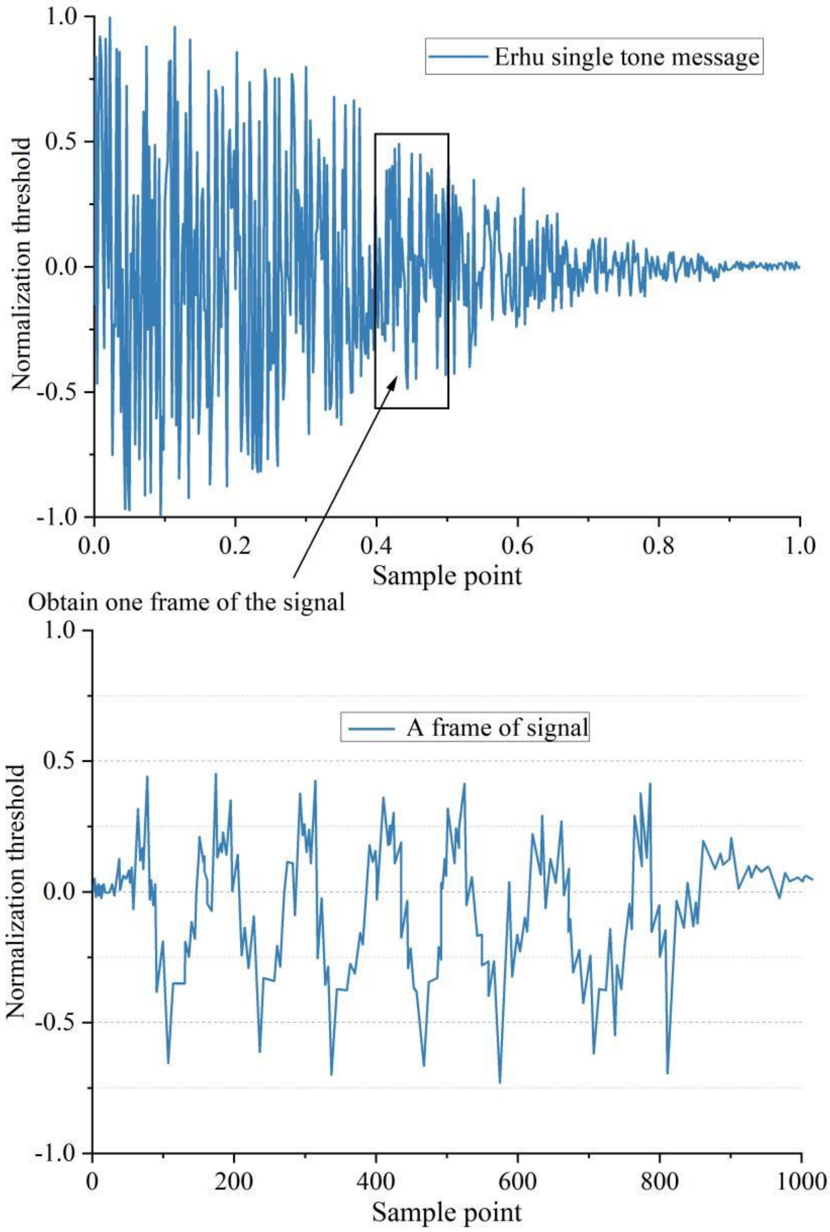

Figure 7: The single-tone signal and the sound signal of one frame of of erhu music

As shown in Figure 7, the single-tone waveform of the erhu consists of three different stages, namely, the initial excitation, steady-state, and decay. When the excitation happens, the amplitude quickly increases, which is due to the transient nature of the bow-string friction. The steady state phase is characterized by the amplitude leveling off with strong periodic oscillations. The decay phase is characterized by a slow decrease in amplitude, which shows the damping of the resonance of the instrument. Zooming into a narrow time interval unveils a higher level of periodic oscillation construction in the waveform.

The figure 8 depicts the autocorrelation function of the musical signal in one of the frames of this erhu.

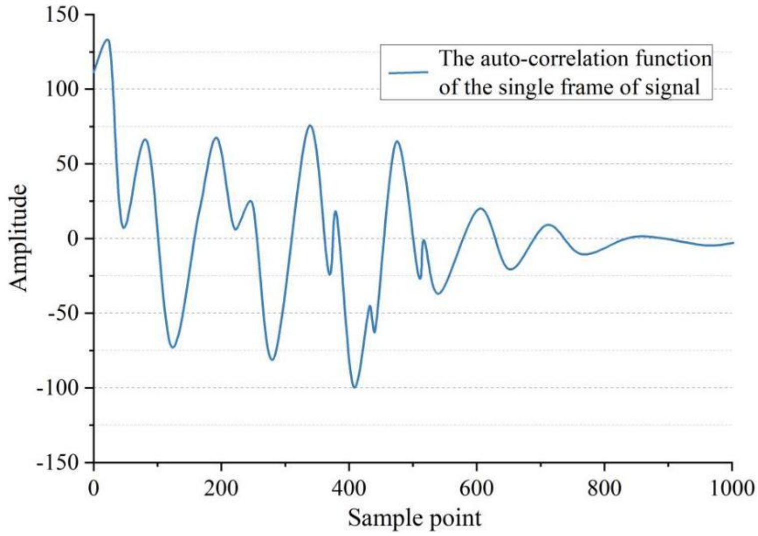

Figure 8: The auto-correlation function of a certain frame of the erhu sound signal

The autocorrelation curve has characteristic periodic peaks, where the peak spacing of the main peaks corresponds to the period of the fundamental frequency. At T0=5ms, the fundamental frequency has a value of 150Hz (f=1/T0). The fundamental frequency period may be accurately determined by identifying the amount of delay that makes the signal most similar to its delayed version. The first significant peak location in the figure indicates the fundamental frequency period and other secondary peaks indicate the periods of harmonic components. The technique addresses noise interference, which is essential to obtain important parameters used in further analysis of the harmonic structure, and gives the guarantee of robustness when extracting timbre features.

3.2.3 Analysis of Classification Results at Different Highest Harmonic Orders

According to the basic detection of the autocorrelation function, timbre features with various maximum harmonic orders were built. The effect of harmonic order on the recognition accuracy of different instruments was considered and compared. An experiment was performed to determine a suitable maximum harmonic order whereby the recognition of Chinese traditional instruments was tested using the harmonic structure-based timbre feature extraction method that had been developed. The recognition accuracy rates of the different instruments at different maximum harmonic orders are presented in the table 3 and figure 9.

Table 3: The recognition accuracy of different maximum harmonic orders

The highest harmonic order	Stringed instruments	Stringed instruments	Percussion instruments	Blowing instruments
3	80.41	81.92	84.71	87.81
4	83.44	84.86	86.62	90.51
5	86.14	87.97	90.28	93.62
6	89.80	93.38	94.97	97.19
7	93.54	97.14	97.95	99.11
8	96.24	94.35	95.68	98.24
9	94.81	89.56	96.34	95.53
10	92.19	86.46	92.19	93.14
11	89.23	83.91	88.29	90.04
12	87.02	81.68	85.62	87.81

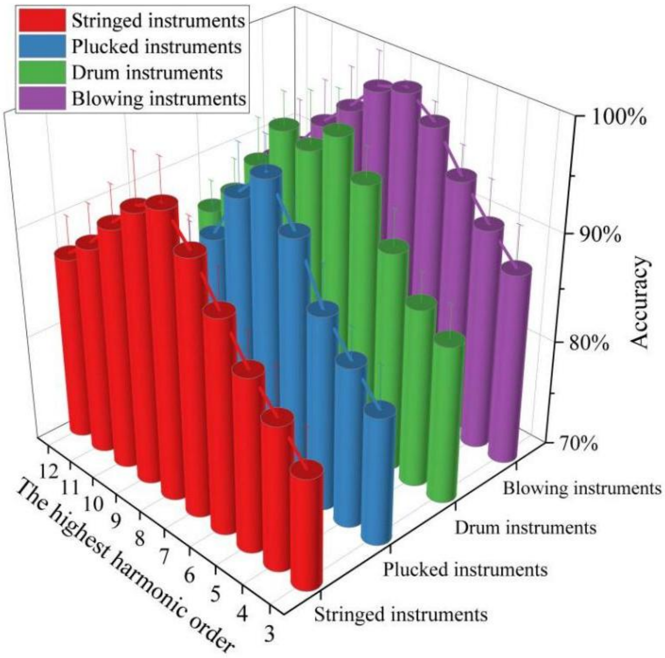

Figure 9: The recognition accuracy of different maximum harmonic orders

The table and image data clearly demonstrate that recognition accuracy does not increase monotonically with the number of harmonics. Conversely, it is possible to identify another type of so-called optimal range where after exceeding this range, the performance will decline. In particular, when the highest harmonic order slowly changes between the third and the seventh order, the recognition accuracy of all four instrumental classes demonstrates a consistent and significant increase. There was an increase in accuracy, i.e., 80.41-93.54 (bowed instruments), 81.92-97.14 (plucked instruments), 84.71-97.95 (percussion instruments) and 87.81-99.11 (wind instruments), reaching almost perfect recognition. This stage of better accuracy implies that adding higher-order harmonic elements better describes the spectral structure of an instrument timbre, which offers more detailed identification data. Nevertheless, once the harmonic order went beyond the 7th, the accuracy started to demonstrate a point of inflection and often decreased. Stringed instruments attained a peak of 96.24 percent accuracy at the 8th harmonic but then fell off, whereas plucked, percussion, and wind instruments hit their peak at the 7th harmonic and then all steadily dropped. Accuracy in all categories drops to less than 90% by the 12 th harmonic, which is lower than levels at fewer harmonics. It is possible that too many harmonics could be used to add too much detail noise or irrelevant spectral data, resulting in obscuring key timbre features and poorer classifier performance. This can be connected to overfitting or signal-to-noise ratio reduction caused by the loss of high-frequency energy.

4 Simulation Study Based on an Agent-Based Music Dissemination Model

Having finished with the technical construction and experimental validation of a system to extract tonal features, classify and identify Chinese traditional instruments, this paper goes on to examine the processes of the diffusion of musical information across the globe through the lens of computational simulations. Building upon the previously proposed agent-based music propagation model, we conduct in-depth investigations into the dynamic propagation behavior of musical information within complex social networks through numerical simulations and network evolution analysis.

4.1 Research on the Stability of Knowledge Dissemination Equilibrium Points

Numerical simulations were conducted using the Runge-Kutta algorithm to validate the theoretical analysis results. Since existing studies lack universal guidelines for parameter values, this paper provides numerical values for each model parameter based on the fundamental reproduction number R_0 and stability conditions.

To validate the stability of the equilibrium point for musical information propagation in Chinese traditional instruments, the parameters are set as $\Lambda=1.2$, $\mu=0.3$, $\alpha=0.4$, $\beta=0.5$, $r=0.5$, and $\theta=0.5$, yielding $R_0=3.5>1$. Numerical simulations using the Runge-Kutta algorithm show the densities of unknowns, propagators, super-spreaders, and immune individuals at steady state, as depicted in Figure 10.

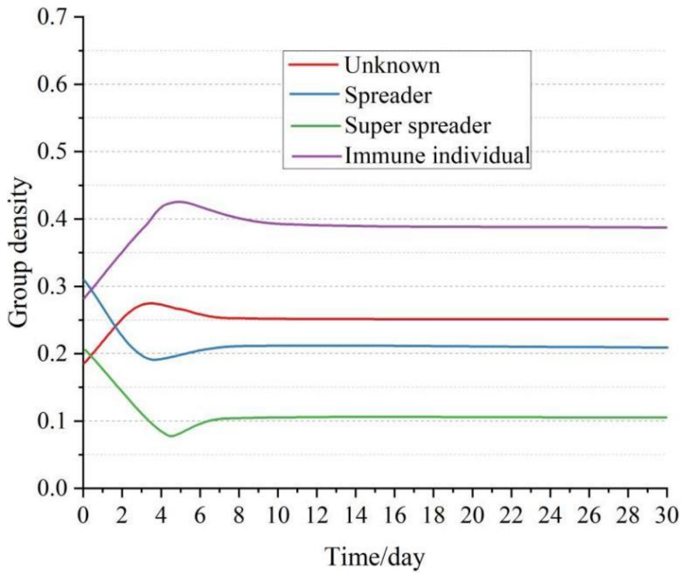

Figure 10: The density of each group of the balance of knowledge dissemination

As shown in Figure 10, the initial density distribution ratio of the unaware, spreaders, super spreaders, and immune groups is 2:3:2:3. After around 8 days of music information transfer, a new steady state equilibrium is achieved with the group ratios changing to 3:2:1:4. Spreader decreases and the densities of Spreader and Super Spreader groups also decrease due to the natural decline of transmission momentum after an original outbreak. The size of immune people grows, as this population keeps on growing over time, which is the manifestation of the so-called herd immunity effect when information saturation is reached. On the whole, the equilibrium point of knowledge dissemination does not change. It has been confirmed that such stability will be present with other initial conditions.

4.2 The Impact of Knowledge Subject Personality Traits at Different Network Layers on Knowledge Dissemination

The stability analysis discussed above shows the convergence properties of the music spread system on a macro level. In order to gain additional insight into the micro-level processes of dissemination dynamics, the next section will discuss how the unique features of individual knowledge agents influence the dissemination process by adding the hierarchical structure of the network. It is possible to divide the network into three levels namely In, Middle and Out, and manipulate the attribute parameters of the nodes in each level to be able to better approximate knowledge diffusion routes in heterogeneous networks which would integrate macro-level stability with micro-level agent behavior.

Within the knowledge network system, connections between knowledge agents and their network partners—both across layers and within layers—are dynamically realized through adjustable probability matrices.

The individual characteristics of knowledge agents are expressed through their personality attribute values within each network layer. In the simulated evolution of the Chinese traditional musical instruments knowledge network, p represents the enthusiasm of knowledge agents in participating in knowledge dissemination (set to p=0.2 in this simulation). When knowledge agents participate in knowledge dissemination with a certain level of willingness, the evolutionary process of the knowledge network within the system unfolds as follows.

4.2.1 Network Knowledge Evolution Process for Different In-Layer Attributes X

Figure 11 illustrates the network knowledge evolution process when the Middle and Out network layer knowledge entities' personality attribute values are both 0.2 (i.e., Y=0.2, Z=0.2), and the In network layer knowledge entity's individual attribute value X is set to 0.2, 0.4, 0.6, and 0.8 respectively.

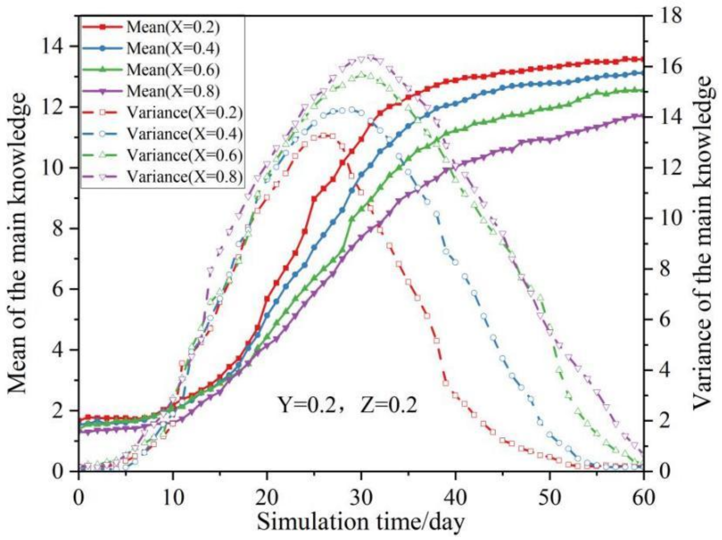

Figure 11: The process of network knowledge evolution occurs of different attribute X

As the attribute value X of knowledge agents in the In network layer increases, the growth rate of the system's average knowledge value slows significantly, and the time required to reach dynamic equilibrium notably lengthens. This occurs because agents in the In layer, being highly connected nodes, exhibit a tendency to carefully evaluate information value when their attribute X (propagation proactivity) is higher. This cautious evaluation, paradoxically, slows the rate of knowledge diffusion.

4.2.2 Network Knowledge Evolution Process under Different Middle Layer Attributes Y

Figure 12 illustrates the network knowledge evolution process when the personality attribute values of knowledge entities in both the In and Out network layers are 0.2 (i.e., X=0.2, Z=0.2), and the individual attribute value Y of knowledge entities in the Middle network layer is set to

0.2, 0.4, 0.6, and 0.8 respectively.

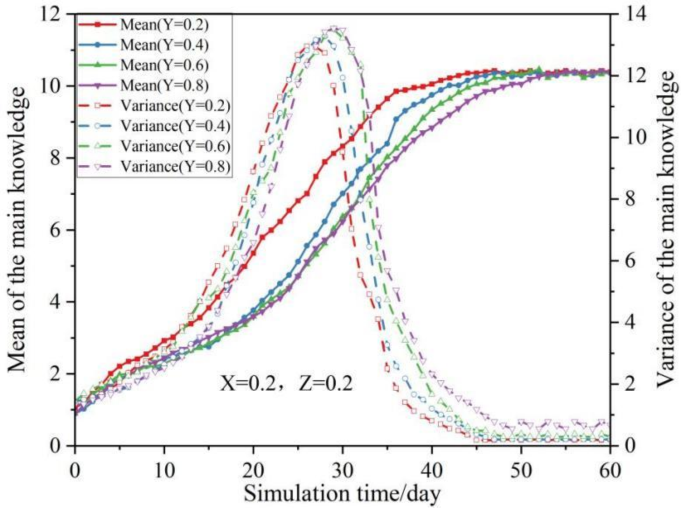

Figure 12: The process of network knowledge evolution occurs of different attribute Y

The individual attributes of knowledge agents at the Middle network layer exert a relatively minor influence on the time required for the knowledge system to reach dynamic equilibrium. However, as the value of individual attribute Y increases, the growth rate of the average knowledge level among network agents shows a noticeable decline. As a bridge group, changes in the attributes of Middle-layer agents primarily affect knowledge absorption intensity rather than the propagation topology, indicating that Middle-layer nodes are more inclined toward “information filtering” than “path reconstruction.”

4.2.3 Network Knowledge Evolution Process under Different Out-layer Attributes Z

Figure 13 illustrates the network knowledge evolution process when the knowledge subject's personality attribute values in both the In and Middle network layers are 0.2 (i.e., X=0.2, Y=0.2), and the individual attribute value Z in the Out network layer is set to 0.2, 0.4, 0.6, and 0.8 respectively.

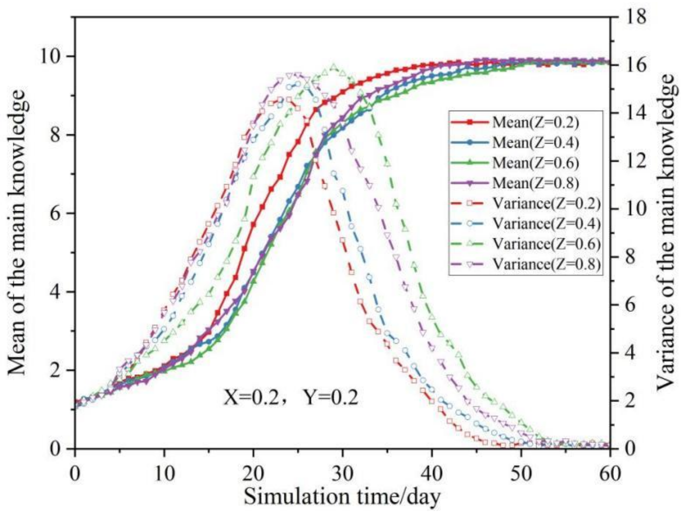

Figure 13: The process of network knowledge evolution occurs of different attribute Z

As the individual attribute value Z of knowledge subjects in the Out network layer increases, the growth rate of the mean subject knowledge exhibits a trend of first decreasing, then increasing, and finally decreasing again. The mean reaches its lowest point when Z=0.6, indicating that the growth rate of the mean subject knowledge displays a discontinuous characteristic. After Z > 0.6, the growth rate rebounds, indicating the existence of a "participation threshold" in the Out layer (general audience). Specifically, when reception willingness reaches an intermediate value, the group tends to fall into "selective numbness" due to information overload.

Overall, within the dissemination of knowledge about Chinese traditional musical instruments, the individual attribute values of knowledge subjects in the In and Out network layers exert a strong influence on the system's average knowledge value and the evolution time of its dynamic equilibrium. Conversely, the individual attribute values of knowledge subjects in the Middle network layer have a relatively weaker impact on both the system's average knowledge value and its dynamic equilibrium evolution time.

5 Conclusion

This study conducted a quantitative analysis and dynamic simulation of the dissemination pathways of Chinese traditional musical instruments on the global music stage. The effectiveness of the suggested model was confirmed by experimental data and simulation results, which also showed important trends in the process of dissemination.

On the basis of conventional spectral attributes (spectral centroid, spread spectrum) addition of new first- and second-order derivative properties which reflect the time dynamics considerably improves the accuracy of classification. The mean F1 score of 12 instruments has been increased by more than 10 percentage points after averaging over 86.29% and 96.42%. Additional research on harmonic structure revealed that setting the highest harmonic to 7 yields 93.54%, 97.14%, 97.95%, and 99.11% recognition rates of bowed, plucked, struck, and blown instruments respectively, indicating the best trade-off between tonal richness and model robustness.

The simulation outcomes using the Runge-Kutta algorithm and the agents suggest that, in case of the basic reproduction number $R_0 = 3.5 > 1$, there is a steady equilibrium point in the transmission system. Population densities stabilise after about eight days (uninfected, infected, super-spreaders and immune ratios are about 3:2:1:4). Network-layered simulations further quantified the influence of nodes with different attributes. When the transmission propensity parameter X of core nodes (In layer) increased from 0.2 to 0.8, the time required for the system to reach dynamic equilibrium significantly lengthened. Meanwhile, the parameter Z of ordinary audiences (Out layer) exhibited a "participation threshold" near 0.6, causing a jump in the growth rate of the system's average knowledge level.

References

[1] Zhang, T. B., & Liu, L. (2018, November). Improving the Soft Power of Chinese Culture and Enhancing the Countermeasures of Cultural Confidence. In 3rd Annual International Conference on Education and Development (ICED 2018) (pp. 178-183). Atlantis Press.

[2] Nur Mutia, R. T., & de Archellie, R. (2023). Reassessing China's soft power in Indonesia: A critical overview on China's cultural soft power. Cogent Arts & Humanities, 10(1), 2178585.

[3] Du, B., & Liang, Y. (2024). The Development and Evolution of Ancient Chinese Musical Instruments. Material science, 6(7), 46-51.

[4] Xu, K. (2021). Recognition and classification model of music genres and Chinese traditional musical instruments based on deep neural networks. Scientific Programming, 2021(1), 2348494.

[5] Ma, Y., & Chen, Y. (2024). Exploring the model of contemporary Chinese ethnic musical instrument improvement mechanisms: Based on grounded theory. Sage Open, 14(1), 21582440241235018.

[6] Ziyin, L. I. U. (2015). On the Inheritance and Innovation of the Cultural Spirit of Chinese National Music. Higher Education of Social Science, 9(5), 64-68.

[7] Shun-Yee, H. (2003). The significance of musical instruments and food utensils in sacrifices of Ancient China. Monumenta serica, 51(1), 1-18.

[8] Zi, W. (2020, March). Practice of Knowledge: Overview of the First Advanced Workshop on the Archaeology of Chinese Music. In 4th International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2020) (pp. 349-353). Atlantis Press.

[9] Wang, G., Chuangprakhon, S., & Theerapan, S. (2023). An investigation of teaching how to create string instruments of the zhuang nationality in guangxi china. International Online Journal of Education and Teaching (IOJET), 10(2), 729-739.

[10] Juan, L., Jirajarupat, P., & Yinghua, Z. (2023). The Transmission of Guqin Musical Instrument Knowledge Literacy and Its Reflection Study in Guizhou Province, China. International Journal of Education and Literacy Studies, 11(2), 22-29.

[11] Huang, Y., Chuangprakhon, S., & Santaveesuk, P. (2024). Preservation and transmission of Shaanxi Guzheng musical instruments: Challenges and strategies for cultural sustainability. International Research Journal of Multidisciplinary Scope, 5(04), 147-158.

[12] Liu, S., & Bhengsri, T. (2024). Historical Perspective on Literacy and Learning Resources of Traditional Chinese Fiddle Musical Instruments in Jiangxi. International Journal of Education and Literacy Studies, 12(1), 260-267.

[13] Li, Z., & Bryan-Kinns, N. (2021). The Cross-Cultural Design in Joint Efforts: Reshaping Traditional Chinese Musical Instruments in Application of Digital Technology. People's Music, 7, 55-60.

[14] Wu, Z. (2024). The Image Transmission and Cultural Heritage of Chinese Musical Instruments. Frontiers in Art Research, 6(1), 44-50.

[15] Liu, L. (2023). Research on the Foreign Communication Path and Strategy of Chinese Folk Music under the Background “The Belt and Road”—Taking Ningbo as an Example. Journal of Social Science and Humanities ISSN, 1811, 1564.

[16] Hong, X., & Wu, Y. H. (2022). The use of AR to preserve and popularize traditional Chinese musical instruments as part of the formation of the tourist attractiveness of the

national art of Guizhou province. Digital Scholarship in the Humanities, 37(2), 426-440.

[17] Guo, W., Huang, Y., Chen, Z., Zhang, Z., Sun, G., Zeng, Q., & Li, X. (2023). The “rebirth” of traditional musical instrument: An interactive installation based on augmented reality and somatosensory technology to empower the exhibition of chimes. Computer Animation and Virtual Worlds, 34(3-4), e2171.

[18] Shen, D., Weng, J., Xu, R., Wang, L., Zhou, J., Bao, D., ... & Li, C. (2025). OrigamiSOUND: a multisensory interactive digital musical instrument for children to experience traditional Chinese musical instrument culture. Universal Access in the Information Society, 1-12.

[19] Lu, F., & Sornyai, P. (2024). Exploring organology and learning techniques of Dongxiao Chinese folk musical instruments in Shandong province. International Journal of Sociologies and Anthropologies Science Reviews, 4(2), 53-66.

[20] Liang, X., Li, Z., Liu, J., Li, W., Zhu, J., & Han, B. (2019, July). Constructing a multimedia Chinese musical instrument database. In Proceedings of the 6th Conference on Sound and Music Technology (CSMT) Revised Selected Papers (pp. 53-60). Singapore: Springer Singapore.

[21] Yang, H., & Sun, G. (2022). Feasibility study on the construction of musical instrument digital inculcating museum in Xinjiang, China. Forest Chemicals Review, 1901-1912.

[22] Jia, Q. (2021). Application of Chinese traditional musical instruments. Scientific and Social Research, 3(1), 80-84.

[23] Frid, E. (2019). Accessible digital musical instruments—a review of musical interfaces in inclusive music practice. Multimodal Technologies and Interaction, 3(3), 57.

[24] Lim, C. K. N., & Zhang, M. (2023). Chinese national music platformisation: A systematic review. Heliyon, 9(11).