Received: 31 July 2021 | Revised: 3 September 2021 | Accepted: 8 October 2021

Cognitive Computation and Systems

DOI: 10.1049/ccs2.12036

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Influence of musical elements on the perception of 'Chinese style' in music

Yu Zhang[1] | **Ziya Zhou**[1] | **Maosong Sun**[1,2]

[1]Department of Music AI and Information Technology, Central Conservatory of Music, Beijing, China

[2]Department of Computer Science and Technology, Tsinghua University, Beijing, China

Correspondence

Maosong Sun, Department of Music AI and Information Technology, Central Conservatory of Music, Beijing, China.
Department of Computer Science and Technology, Tsinghua University, Beijing, China.
Email: sms@tsinghua.edu.cn

Funding information

National Office for Philosophy and Social Sciences, Grant/Award Number: 21ZD19

Abstract

Recently Chinese music has been regarded as an independent music school. However, the definition of Chinese music has been an abstract and subjective concept. That makes music information retrieval (MIR) tasks hard to perform on Chinese music. Previous musicological studies have explained how musical elements like melody and instruments shape a certain kind of music genre including Chinese music, but the findings cannot be directly applied to relevant MIR tasks towards large-scale users and real-world problems. In this study, a pipeline of performing a perceptual survey is designed to explore how different musical elements influence people's perception of 'Chinese style' in music. Participants with various backgrounds were presented with categorised music excerpts performed in the Erhu or violin and then gave 'Chinese style' ratings. Statistical analysis indicates that music content contributes more than instruments. Results were compared between musicians and non-musicians. Subsequently, a supplementary automatic music classification experiment is conducted in comparison with the survey results to discuss the authors' choice of stimuli in the survey and similarities between computer auditory and human perception. In general, the results in this study can be useful for MIR tasks, such as understanding, representation, and recommendation of Chinese music.

KEYWORDS

automatic music style classification, Chinese music style, music information retrieval, music perception

1 | INTRODUCTION

In recent years, Chinese music tends to be regarded as an independent music school different from other existing categories, including the Vienna Classical Music School, Russian Folk Music School, and Venetian Music School [1–6]. With a long-standing history of Chinese civilisation from the Xia Dynasty to the present, incorporating the culture of diverse ethnic groups, Chinese music blooms into a broad and pluralistic system. Due to the complexity of Chinese music, whose evolution can be informed by notions of orientalism, exoticism, globalisation, transculturation, and hybridity [7], it is hard and unrealistic to define 'Chinese Music School' based on characteristics of those well-developed music schools or some regular patterns. For example, one of the notions accepted extensively is that originated from different culture and aesthetic meaning, Chinese traditional music greatly differs from the harmony-based Western music tradition [8]. A set of algorithms, systems, and tools have been invented and developed for the analysis, extraction, and representation of musical concepts in Western tradition [9]. However, the uncertainty of 'Chinese music' and its music style leaves a great challenge of quantitatively computing Chinese music.

Generally, music genre (or style), a perceptually abstract musical concept that is shaped by concrete musical elements such as rhythm, tempo, melody, instruments, playing techniques, and music structure. A lot of research on music genre or music style have already revealed that melody (music content) plays an important role in shaping a certain kind of music style [10–13]. For example, the Chinese pentatonic scale '宫(*C*)商(*D*)角(*E*)徵(*G*)羽(*A*)' is commonly used as the melody mode in many Chinese music. Also, some instruments have intrinsically national and ethnic characteristics. For example, The Erhu is a Chinese traditional bowed-string instrument with over 1000 years' history and shakuhachi is a Japanese and ancient Chinese longitudinal and end-blown flute made of bamboo. All of the musicological research on music genre is based on the perspectives of historical evolution, instead of generalisation of practical scenario, where music genre is also a very important component in music information processing [14]. Therefore, previous findings may not be applicable to real-world scenarios based on large-scale users. For example, automatic music genre classification (MGC) [15–28] is a significant task in music information retrieval (MIR), helpful for applications like music recommendation systems. However, from the perspective of musicology, music genres are subjective concepts, not having an absolute and permanent definition. Views of the genre of an excerpt may vary from Jazz to Blues due to different perceptual processes, no matter for domain experts or ordinary people [29–31], not to mention their diverse cultural backgrounds. That means a single or even multi-genre label attached to a piece of music is far from being able to represent its abstract musical characteristics. In other words, to better utilise music genre information in music representation, we need to quantitatively disentangle [32–34] concrete musical elements that are easier to extract from unstructured music files based on extensive investigation and survey on people's listening process as well as machine learning experiments.

Lots of previous studies [15–28] focus on improving the classification accuracy by technically advancing the model, but few of them have investigated from the perceptual aspects on quantitatively measuring how different musical elements like melody, instruments, or playing techniques influence people's impression of a certain music style, especially on Chinese music. As numerous literature discussed, it seems to be a common concept that melody plays a more important role than timbre in musical genres. However, the definitions of melody and timbre were not clear and the conclusion might not be robust. For example, possible variables during the performance could be part of the melody or instrument characteristics, including tempo, rhythm, dynamics and performance techniques. In this study, we design a perceptual experiment aiming to investigate how music content and instruments influence people's impression of the Chinese music style. It should be noticed that the variable 'music content' refers to the organisation of melody information that can be specified in musical scores, including tonality, tempo and rhythm. Two categories of musical excerpts, *Melodic* and *Etudes*, were chosen to represent Chinese and Western music, respectively (see details in Section 3.1). *Atonal* music was also included to obtain people's impression of Chinese modern musical works. The Erhu and violin were chosen as instrument variables since they are representative Chinese and Western instruments in the stringed instrument families. Musical scores for Erhu and violin performances were unified from the same excerpt. Although playing techniques are commonly referred to be instrument characteristics, they sometimes involve pitch shifts, and the same playing technique might vary from one performer to another. In this study, to disentangle the factors contributing to people's perception of music, the playing techniques were removed during recording so that the music content could be aligned between Erhu and violin performances. Furthermore, we conducted an automatic classification experiment on audio samples used as listening materials of participants, in comparison with their preference reflected in the perceptual experiment. Here is the pipeline of our work:

Step 1 Choose Erhu and violin as the experimental instruments and determine the music excerpts

Step 2 Record excerpts and unify them into the same range of loudness level

Step 3 Design and publish an online questionnaire for perceptual experiments

Step 4 Collect the results and perform statistical analysis

Step 5 Extract acoustics features from used excerpts and conduct the automatic music classification experiment

Details of each step will be discussed in the following sections. There are several contributions made by this work:

- We designed a perceptual experiment to explore how music content and instruments may influence the perception of 'Chinese style' in music. Performance techniques were excluded to disentangle the instrument variables from music content. Musical scores for Erhu and violin performances were unified.
- A WeChat Mini programme was developed to publish the questionnaire and collect a large number of responses from various groups. We have designed and implemented some technical details to ensure valid responses (see Appendix I). The methods of this experiment can be applied in further research on related issues.
- Participants cover a wide range in terms of age, education and music studies. The questionnaire results reflect people's understandings of Chinese music in the contemporary cultural background. This could be applicable in music recommendation systems and be valuable to musicians in composing Chinese music.
- We also discussed our choice of stimuli and similarities between computer auditory and human perception by conducting an automatic music classification experiment on all excerpts and comparing its results with the perceptual experiment. It shows the excerpts used as stimuli are reasonable. The result also reveals some characteristics of each category of excerpts, which corresponds with findings from some previous relevant qualitative research.

This study is organised in this way: In Section 2, we summarise the previous work on Chinese music, music style classification and music perception. In Section 3, we detail the whole process of the perceptual experiment. In Section 4, we present the statistical analysis of the survey results. In Section 5, we introduce findings of our supplementary music style classification experiment. To clarify our goal, we must point out that this research is not on the purpose of clearly defining 'Chinese Music School' or 'Chinese style' in music, but mainly focus on surveying from a perceptual perspective on how musical elements, such as music content and instrument, influence people's preference on 'Chinese style' in music and producing some findings that are helpful for music information processing and retrieval of Chinese music.

2 | RELATED WORK

2.1 | Chinese music and its computation

The idea that regarding Chinese music as an independent music school has been discussed since the 1990s [1], and the concept of 'Chinese Music School' (中国乐派) was first introduced in 2015 [2]. Chinese Music School should be established based on both music ontology and culture environment, encompassing Royalty Court Music (宫廷音乐) in *Pre-Qin* Period, *Jiyue* (中古伎乐) from *Qin* dynasty to *Five Dynasties and Ten Kingdoms* Period, and folk music from *Song* to *Qing* dynasty [3].

Specifically on computational methods and knowledge representation of Chinese music, Tian et al. [35] examine existing metadata standards for describing music-related information in the context of Chinese music tradition and define new notation systems for Chinese music. Some other studies focus on Jingju, one of the Chinese traditional music forms. Repetto et al. [36] presented the first collection of machine-readable scores for the study of Jingju singing, including 92 scores covering 897 melodic lines accompanied by their metadata and curated annotations per score and melodic line. Based on that, Gong et al. [37] extend the dataset for some evaluation research, Repetto et al. [38] present a quantitative analysis of the relationship between linguistic tones and melody in Jingju, and Zhang et al. [39] propose a novel approach to study the expressive functions of *banshi* (some rhythmic devices used in Jingju) by applying text analysis techniques on lyrics. In a word, there is some MIR-related research on representing and quantitatively evaluating certain types of Chinese music but few on technically interpreting the characteristics of Chinese music style.

2.2 | Automatic music genre (style) classification

Plenty of research on MGC focusses on music content, including audio and symbolic features processing. As for audio, Tzanetakis et al. [15] propose a set of features for representing texture, instrumentation, rhythmic structure, and strength for this task. Baniya et al. [16] propose an extreme learning machine (ELM) with bagging and incorporate timbral texture and rhythmic content features. Ghosal et al. [17] train a convolutional long-short term memory-based neural networks (CNN-LSTM) model on extracting a diverse set of spectral and rhythmic features and a transfer learning model on classifying the genre. Liu et al. [18] exploit the spectrograms of audios and develop a novel CNN architecture suitable for processing long-contextual information on several benchmarks, including GTZAN [15], Ballroom [20], and Extended Ballroom [21]. As for symbolic data such as MIDI files and MusicXML, Basili et al. [22] derive features from the MIDI and uses a set of unsupervised machine learning algorithms such as decision-tree and Bayesian. McKay et al. [23] extract 109 musical features based on instrumentation, texture, rhythm, dynamics, pitch statistics, melody, and chords from MIDI files and perform the feedforward neural networks and k-nearest neighbour hierarchically using different sets of features at different levels. Cataltepe et al. [24] prove that the combination of MIDI files and audio features from MIDI improves accuracy for MIDI MGC. Ferraro et al. [25] apply SIA and P2 algorithms on a large and diverse corpus containing over 40,000 MIDI files.

In recent years, to enrich the music genre representation, some research incorporates other relevant multimodal music information like artists and albums rather than merely content-based approaches. Oramas et al. [26] present MuMu, a new multimodal dataset of more than 31k albums classified into 250 genre classes, and propose an approach for multi-label genre classification based on the combination of feature embeddings learnt with deep learning models. Bogdanov et al. [27] introduce the AcousticBrainz Genre Dataset, a large-scale collection of hierarchical multi-label genre annotations from different metadata sources, exploring how the same music pieces are annotated differently by different communities following their genre taxonomies. Epure et al. [28] study the feasibility of obtaining relevant cross-lingual, culture-specific music genre annotations only based on language-specific semantic representations and show that unsupervised cross-lingual music genre annotation is feasible with high accuracy, especially when combining both types of representations. Lots of relevant datasets are constructed based on annotation of several domain experts. This is problematic because music genre is a subjective concept and the labels cannot represent people's perception on it broadly. In this way, only to improve the classification accuracy is far away from meeting the requirements of MIR applications. Even though most studies of MGC are trying to find more diverse representation of music, they still focus on improving the accuracy of the MGC task rather than effectively representing the music genre as one of the abstract and subjective musical characteristics from perceptual perspectives.

2.3 | Music perception studies

Many studies have investigated the influence of the melody and the timbre changes on music perception [40–43]. The findings in Ref. [40] imply that timbre change, along with pitch and rhythm, strongly affects the recognition of excerpts in Western contemporary and tonal music, with the impact being stronger for musicians than for non-musicians with tonal materials. According to Ref. [41], instrumentation and timbral variations also influence the perception of musical similarity. On the other hand, the research study published in Ref. [42] indicates that musicians always chose melody and harmonic accompaniment over instrumentation when identifying excerpts, but non-musicians did the opposite. Novello et al. [43] find that the degree of impact on participants' ranking on music similarity is hierarchical (genre > tempo > timbre). Other studies [44, 45] further discuss the effects of timbre and other parameters, such as pitch and tempo, on memory for music. Cross-cultural musical understandings also influence musical memory as described in Ref. [46].

Most previous studies on music perception have been focussing on music memory and recognition, and few of them did research on subjective identification of music genres. Inesta et al. [47] conducted a ground-truth experiment on melody genre recognition in the absence of timbre, finding that when presented with a timbre-less segment, individuals can distinguish between well-established genres such as classical and jazz music, but they cannot further explore the underlying musical elements that may contribute to perceptual genre classification. Furthermore, all referenced papers concentrate on Western music and its classification in established Western genres. Their methodologies are limited and could not be directly applied to the perception of 'Chinese style'. Thus, we propose an original methodology to obtain participants' 'Chinese style' perception of music as described in Section 3.

3 | METHODS

To disentangle the influence of music content and the playing instruments on listening test results, the experiment was designed from the following perspectives.

Instruments. We chose the Erhu, one of the Chinese traditional instruments, and the violin, one of the Western classical instruments, as variables in our perceptual experiment, considering their tuning, range, and sounding principles (they are both bowed-string instruments).

Performance. To minimise the performance differences of Erhu and violin performers, all unique playing techniques (including vibratos and trills that sound differently on the Erhu and the violin) were excluded. Instead, we simply kept common bow techniques that could hardly be identified between the two instruments even by professionals. Both performers were asked to perform exactly according to the musical sheets with tempos and beats. Personal performance styles were minimised as well since they could also affect the results.

Recording Apparatus. Each excerpt was performed in the same recording environment. The digital audio workstation was Logic Pro X 2, and the audio interface was the YAMAHA UR22C. The sound from the performance was captured using the Rode M5 cardioid microphone. We placed the microphone near the sound hole of the Erhu and the violin to ensure the concordance of the pick-up quality. To reduce the sharpness of the tone in high-frequency string instruments, the volume level was balanced and fine-tuned for all recorded materials (especially between the Erhu and the violin instruments). The loudness standard was based on the EBU measurement method. The output volume after fine-tune was lower than −3 db. The Short Term Loudness[1] was between −15 and −24 LUFS and Integrated Loudness[2] was between −15.8 and −22.5 LUFS. We have also examined all excerpts by human to ensure that the volume levels were consistent.

3.1 | Stimuli

A variety of excerpts were selected to represent Chinese and Western music, considering participants with different backgrounds and musical understandings of 'Chinese style'. To avoid the individual discrepancies of prior knowledge, well-known melodies were excluded. 116 excerpts were chosen based on five categories: *Chinese-Melodic*, *Chinese-Etude*, *Western-Melodic*, *Western-Etude*, and *Atonal*. The selected excerpts span a wide range of tempos (from 40 to 100 bpm) and emotions (happy and sad). They reside in the common range of the Erhu and the violin. The full stimuli list is provided in Appendix II. *Chinese-Melodic* excerpts covered a wide range from the ancient (Yuan Dynasty) to diverse periods of the 20th century (Liu Tianhua, Wang Yi, and Su Anguo). *Western-melodic* excerpts consist of classical works composed by Charles Dancla, Mozart, and Seitz. Etudes are written for the exercise of playing techniques. It should be noticed that only common bow techniques were kept in the selected excerpts to ensure performance consistency as described above. Compared to melodic excerpts, most etudes consist of simple musical scales and the melody lines are more monotonous. Ten Erhu Etudes composed by Liu Tianhua were chosen to represent *Chinese-Etude*. As for *Western-Etude*, 7 excerpts were selected from *Kayser violin etudes* and four excerpts were selected from *Perception Motion* composed by Ottokar Novacek.

Atonal music written by Chinese composers represents the combination of Western composition techniques and Chinese traditional tunes. Six excerpts were chosen from *Night Scene* written by Sang tong in 1947, and *Erhu Rhapsody No. Two* by Wang Jianmin in 2001. Atonal music is composed without a hierarchical tonal structure and they are characterised by unpredictable combinations of notes, rhythms, and chord progressions. As the first Chinese atonal music work, *Night Scene* adopted composition techniques of serialism and the 12-tone row created by Arnold Schonberg. Pentatonic chords were embedded with small weight to combine atonality with Chinese traditional tunes [48, 49]. *Erhu Rhapsody No. Two* was based on Hunan Huagu Opera and extended by the artificial musical scales [50].

[1] Short Term Loudness measures the loudness of the past 3 s.
[2] Integrated Loudness measures the loudness over the whole track.

3.2 | Procedure

To collect as many results from people with various backgrounds, we developed a WeChat Mini programme[3] since it has a large user base and short development cycle. Before listening to the songs, subjects were asked to fill in the questionnaire that collects background information, including age group, highest education, residence time in China, and music education level. N/A was provided for participants who did not prefer to disclose personal information. Then, they listened to 12 excerpts, and after each one, they were asked to rate using a 5-point Likert scale regarding the following question: '*To what degree do you think it's 'Chinese style' music? Please rate from 1 to 5; 1 means it's absolutely not Chinese style, and 5 means it's completely Chinese style.*' Since the *off-site* experiment was relatively out of control, a series of measures were designed and carried out to ensure valid responses. Detailed information is provided in Appendix I.

The 12 excerpts consist of a 2 + 9 + 1 structure that the subjects were not informed of. The first two songs served as warm-up experiments that were not considered in the statistical analysis. The warm-up songs were randomly chosen from the two categories: Chinese melodic excerpts played in the Erhu, and Western melodic excerpts played in the violin. They are comparatively easy to rate the degree of 'Chinese style'. The following presented nine excerpts were chosen from the following categories: *Chinese-Etude-Erhu*, *Chinese-Etude-Violin*, *Western-Etude-Erhu*, *Western-Etude-Violin*, *Chinese-Melodic-Erhu*, *Chinese-Melodic-Violin*, *Western-Melodic-Erhu*, *Western-Melodic-Violin*, *Atonal-Erhu and Atonal-Violin.* Only one excerpt was chosen from the *Atonal-Erhu* and *Atonal-Violin* because of its small percentage. The playing order was in random. One questionnaire did not contain the same excerpts played in different instruments. The last excerpt repeated the first one to obtain the stability of the experiment.

In total 601 valid responses were collected. As shown in Table 1, 601 participants' ratings were for the eight categories, that is *Chinese-Etude*, *Western-Etude*, *Chinese-Melodic* and *Western-Melodic*, with each played in the Erhu and the violin, respectively. The number of responses on *Atonal-Erhu* was 378 and *Atonal-violin* was 223. To eliminate the individual discrepancies of excerpts, the statistical analysis was based on the categories rather than the excerpts.

4 | RESULTS

4.1 | Participants

There were over 1000 visits to the programme in the 10 days after the release and 601 complete responses were obtained. The online survey's participants showed a wide range of backgrounds. As shown in Figure 1, participants were classified based on music education level, age group, residence time in China, and highest education level. The majority in each category were amateurs trained with teachers, age 18–25, residence time in China for longer than 15 years and undergraduate, respectively. The music education level category was roughly even-distributed according to the four groups: (1) people with no music studies, (2) amateurs (self-trained), (3) amateurs (trained with teachers) and (4) musicians who have been educated in the professional musical academy.

[3] https://developers.weixin.qq.com/miniprogram/dev/framework/

TABLE 1 Number of responses on musical excerpts

	Playing instruments	
Category	**Erhu**	**Violin**
Chinese-Etude	601	601
Western-Etude	601	601
Chinese-Melodic	601	601
Western-Melodic	601	601
Atonal	378	223

4.2 | Reliability and validity

In this experiment, the test-retest reliability, also known as the stability coefficient, measures the subject's stability at the beginning and end of the listening survey. As described in Section 3, at the end of the listening test, each participant listened to the music that was already played at the start of the test. The correlation coefficient was calculated based on the answers from all participants between the two ratings of the same excerpts. The listening material was randomly chosen so that the repeated excerpts were different for participants. The correlation coefficient is 0.812, which demonstrates the good reliability of the experiment.

The validity test was carried out through factor analysis, subjecting to two conditions: The KMO (Kaiser-Meyer-Olkin) value needs to be greater than 0.6, and the Bartlett spherical p-value is less than 0.05. The validity test results (see Table 2) show that the KMO value was 0.752, and the Bartlett spherical p-value was 0.01. Hence, the results are suitable for factor analysis.

4.3 | Factor analysis

Factor analysis showed that the cumulative percentage of three factors achieved 54%, of which Factor 1 and 2 were the principal factors that can explain 42%. *Chinese-Etude-Violin* was attributed to an additional Factor 3, indicating that they shared less commonality with other groups. Factor 2D plot (see Figure 2) presented that the categories were generally grouped according to the *Chinese-* (in red) and *Western-* (in blue) categories. The result suggests that participants classified 'Chinese style' music according to the playing content instead of the performed instruments.

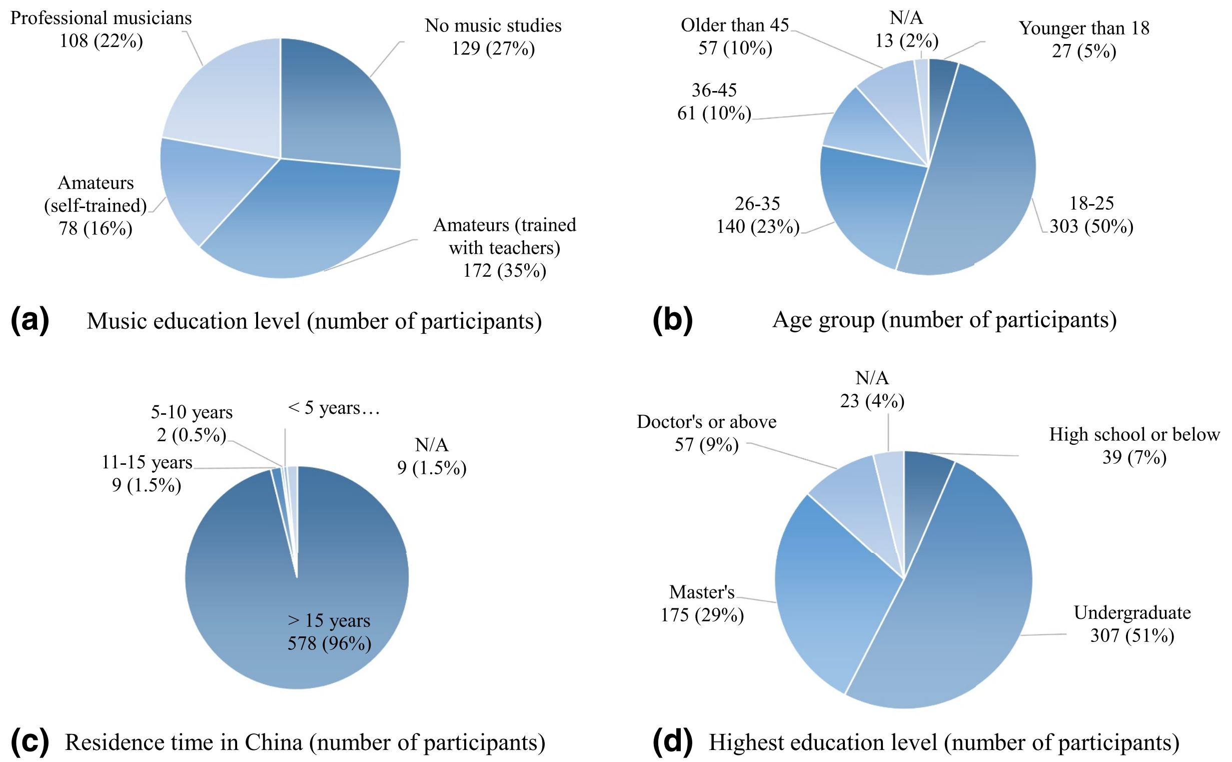


FIGURE 1 Distribution of participants in terms of (a) Music education level, (b) Age group, (c) Residence time in China, and (d) Highest education level

TABLE 2 Validity test results

	Factor loading		
Musical excerpts	**Factor 1**	**Factor 2**	**Factor 3**
Western-Melodic-Erhu	**0.662**	0.033	0.189
Western-Melodic-Violin	**0.623**	−0.285	−0.046
Western-Etude-Erhu	**0.709**	−0.05	0.092
Western-Etude-Violin	**0.755**	−0.114	−0.064
Atonal	**0.612**	0.134	−0.354
Chinese-Melodic-Erhu	−0.158	**0.7**	0.144
Chinese-Melodic-Violin	−0.184	**0.652**	0.118
Chinese-Etude-Erhu	0.271	**0.612**	-0.243
Chinese-Etude-Violin	0.103	0.122	**0.886**
Accumulative contribution rate %	26.849%	42.627%	54.349%
KMO value		0.752	
Barlett spherical p value		0.01	

Note: For example, 'Western-Etude-Erhu' represents Western etudes performed with the Erhu. The significance bold is the Barlett spherical p value listed in the table, i.e., 0.01.

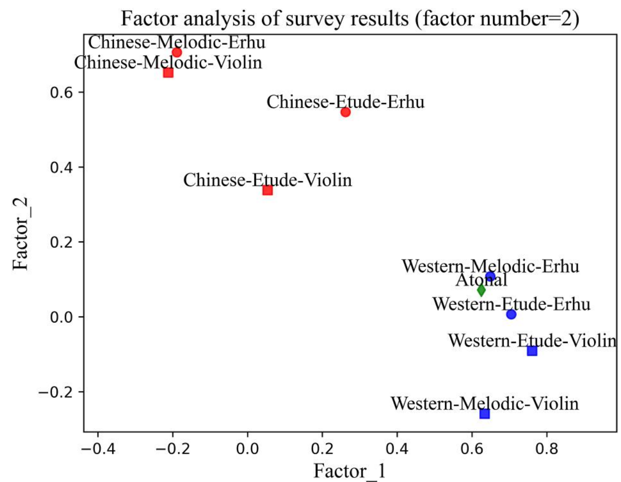


FIGURE 2 Factor analysis of survey results (factor number = 2). The red colour represents Chinese-excerpts. Blue represents Western-excerpts. Circle dots are for excerpts performed in the Erhu instrument, while square dots are for those performed in the violin. The green one is for all atonal excerpts

In addition, it was observed that *Chinese-Etude* excerpts were difficult to classify as 'Chinese style' or not. Possible reasons could be that most people are not familiar with Erhu etudes that consist of scales without melodic lines. Another interesting aspect of this figure is that among *Western* categories, survey results of excerpts played with the same instrument (circle dots for the Erhu and square dots for the violin) tend to be closer, indicating that the instrument affects people's ratings locally within Western categories. Atonal music (in green) was close to *Western-*categories. This could suggest that the people's perception of atonal music was similar to Western music. Since the atonal excerpts (in green) were randomly selected in perceptual experiments without distinguishing instruments, they were not listed in factor analysis. Further discussion on atonal music will be in the following sections.

4.4 | Influence of music content and the playing instrument on 'Chinese style' ratings

The music materials were classified according to music content and playing instruments as shown in Table 1. The statistical analysis of different playing content and instruments was carried out to explore which is more significant on the 'Chinese style' rating. Since the ANOVA normality assumptions did not apply in the survey results, the Kruskal–Wallis test by ranks was used to determine whether a statistically significant difference exists between the medians of multiple independent groups.

As shown in Figure 3, the results present significant differences ($p < 0.01$) in most cases between different instruments and musical categories, so the median differences should be compared. The median differences between instruments under each category was zero or one, much smaller than those between *Chinese* and *Western* categories (e.g. median difference between *Chinese-Melodic-Erhu* and *Western-melodic-Erhu* was three). This observation is highly consistent with the conclusion revealed in factor analysis (Figure 2), where excerpts were generally grouped according to *Chinese* (in red) and *Western* (in blue) categories. The highest rating of 'Chinese style' was observed in *Chinese-Melodic* categories, and the lowest was observed in *Western-Etude* categories. The median and distribution of atonal was close to *Western-Melodic* categories. This manifests that it's hard for people to correspond with atonal music written by Chinese composers to 'Chinese style', although Chinese traditional melody was adopted and combined with Western composition techniques. The median ratings of Erhu-played excerpts are statistically higher than violin-played ones. This indicates that even without playing techniques, the timbre difference between the two instruments are distinguishable and the Erhu gives an impression of Chinese style to the majority people.

When it comes to the results from different groups of participants, no statistically significant bias was observed among groups in terms of the age and highest education levels. Due to the extreme distribution of participants by residence time in China (96% longer than 15 years), the cross-cultural effects on musical understandings could not be analysed. People with different music education levels showed discrepancies. First, the IQR (Interquartile Range, 25th to the 75th percentile) of ratings by people with professional music education (shown in Figure 4a) tends to be smaller than those with no music studies (shown in Figure 4b). This demonstrates that ratings on 'Chinese style' classification by professional musicians are more concentrated. Second, it can be seen that the median differences between *Chinese-Melodic* and *Western-Melodic*, *Chinese-Etude* and *Western-Etude* are greater in survey ratings from professional musicians than those from non-musicians. In other words, professional musicians showed much higher abilities identifying music content compared with non-musicians.

Moreover, the statistical differences between instruments are not significant for all categories in Figure 4b non-musicians. However for (a) musicians, significant differences were observed between instruments for *Atonal*, *Western-Melodic* and *Chinese-Melodic*. This illustrates that musicians are more sensitive to instruments even played without performance techniques and they are able to distinguish them by rating the 'Chinese style'.

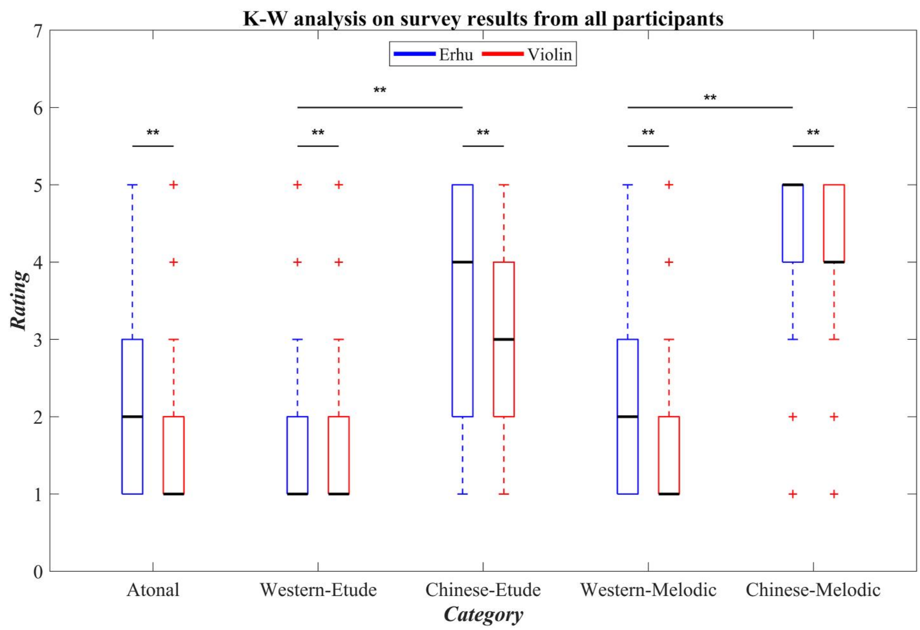


FIGURE 3 The Kruskal–Wallis test results of all participants' ratings on excerpts in terms of five musical categories and Erhu (in blue) versus violin (in red) instruments. ** represents $p < 0.01$

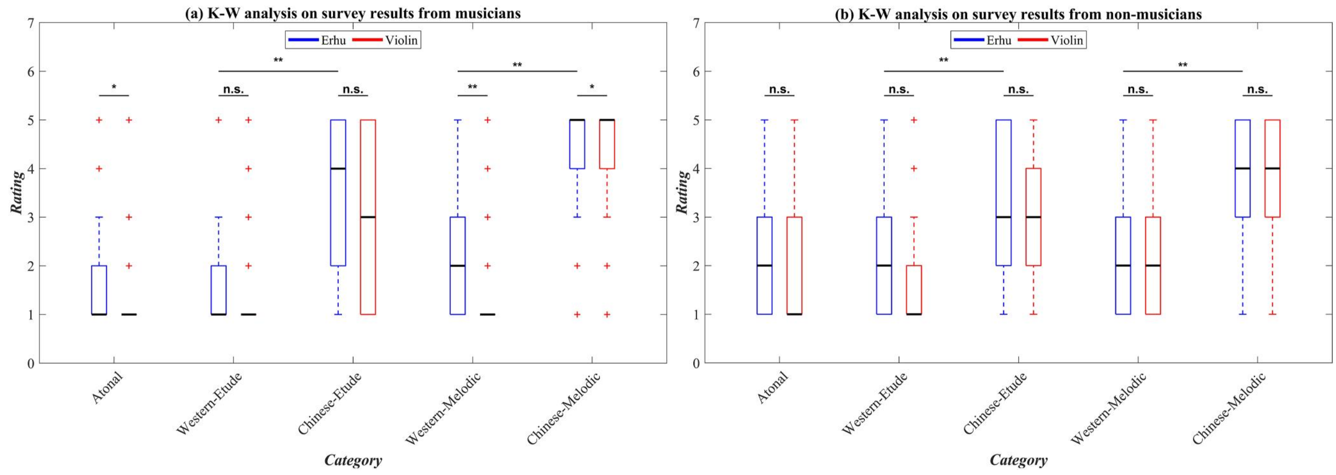


FIGURE 4 Kruskal–Wallis test results of (a) musicians and (b) non-musicians (no music studies) under different categories and instruments. * represents $p < 0.05$ and ** represents $p < 0.01$. n.s. is used for not significant

These results could be helpful in customising recommendation list for different groups of people.

Despite the discrepancies between musicians and non-musicians, common conclusions can be drawn from Figure 4. The difference between *Chinese* and *Western* music categories is much more significant than that between the Erhu and the violin under the same category. In general, *Chinese-Etude* excerpts were difficult to be rated on the degree of 'Chinese style' for all participants.

4.5 | Correlation matrix

Correlation analysis was carried out to further explore the relationship between music categories (see Table 3). Since the survey ratings are based on a 5-point Likert scale (discrete ordinal data), and they are not aligned with normal distribution, Spearman's rank correlation coefficient was utilised to calculate the correlation matrix of all categories. The coefficients in bold ($p < 0.05$) represent the statistically significant correlations between the corresponding categories. A closer inspection of the table shows that the categories with similar music content have a stronger positive correlation regardless of playing instruments. For example, *Western-Etude-Erhu(1)* has a stronger positive correlation with *Western-Etude-Violin(2)*, followed by *Western-Melodic-Erhu(3)* and *Western-Melodic-Violin(4)*. Another example is that a strong positive correlation was found between *Chinese-Melodic-Erhu(7)* and *Chinese-Melodic-Violin(8)*. In contrast, no significant correlation was observed between *Chinese-Melodic-Erhu(7)* and *Western-Melodic-Erhu(3)*.

A negative correlation $R_s(601) = -0.152$, $p < 0.05$ was observed between *Western-Melodic-Violin(4)* and *Chinese-Melodic-Violin(8)*. With a large number of pairs ($N = 601$), the negative correlation (−0.152) was a statistically representative of the population [51]. Moreover, no significant correlations were found between atonal excerpts and other categories (see Table A1 in Appendix IV for the correlation matrix of atonal music).

4.6 | Conclusion of survey results

The critical goal of this survey is to investigate the influence of music content and the playing instrument on 'Chinese style' ratings. The results present high test-retest reliability and validity. Factor analysis demonstrates that the participants' evaluations on music mostly rely on the music content rather than the playing instrument. *Chinese-Etude* excerpts are relatively difficult to rate in terms of 'Chinese style' since people are less familiar with them. Further statistical analysis on the survey results suggests that the rating difference between *Chinese* and *Western* music is much more significant than that between the Erhu and the violin. Distribution of ratings on atonal music is close to Western music. The significant difference between the Erhu and the violin suggests that even without playing techniques, the Erhu is distinguishable compared to the violin and it gives an impression of Chinese style to the majority people. Results also indicate that musicians' ratings are less fluctuated than non-musicians'. It suggests that musicians have higher sensibility in identifying music content and instruments.

Furthermore, the correlation matrix analysis strengthens the idea that excerpts with similar playing content have a stronger positive correlation than those performed on the same instrument. Negative correlations were observed between Chinese and Western excerpts, even if they were performed on the same instrument. Atonal excerpts are less correlated with other groups overall.

5 | SUPPLEMENTARY EXPERIMENT

In music information processing, acoustic features are commonly extracted to represent a certain excerpt for computer audition. Plenty of research [15–21] focusses on extracting acoustics features effectively so that by processing these features, the machine can 'hear' and 'understand' music as human beings do. For example, the mel spectrogram, which evolves from the general spectrogram, improves with the mel

TABLE 3 Correlation matrix of 'Chinese style' ratings on *Chinese*- and *Western*-categories (specifying music content and playing instruments)

Category	Western-Etude-Erhu(1)	Western-Etude-Violin(2)	Western-Melodic-Erhu(3)	Western-Melodic-Violin(4)	Chinese-Etude-Erhu(5)	Chinese-Etude-Violin(6)	Chinese-Melodic-Erhu(7)	Chinese-Melodic-Violin(8)
Western-Etude-Erhu(1)	1	**0.393**	**0.349**	**0.268**	**0.142**	0.075	**-0.084**	**-0.12**
Western-Etude-Violin(2)	**0.393**	1	**0.302**	**0.388**	0.041	0.026	**-0.132**	-0.078
Western-Melodic-Erhu(3)	**0.349**	**0.302**	1	**0.244**	**0.083**	0.045	-0.01	-0.073
Western-Melodic-Violin(4)	**0.268**	**0.388**	**0.244**	1	-0.008	-0.025	**-0.099**	**-0.152**
Chinese-Etude-Erhu(5)	**0.142**	0.041	**0.083**	-0.008	1	0.026	**0.14**	**0.124**
Chinese-Etude-Violin(6)	0.075	0.026	0.045	-0.025	0.026	1	**0.095**	0.068
Chinese-Melodic-Erhu(7)	**-0.084**	**-0.132**	-0.01	**-0.099**	**0.14**	**0.095**	1	**0.262**
Chinese-Melodic-Violin(8)	**-0.12**	-0.078	-0.073	**-0.152**	**0.124**	0.068	**0.262**	1

Note: Correlation coefficients in bold where $p < 0.05$. $N = 601$.

scale [52] according to the human perceptual process instead of the linear Hertz scale. As a result, mel spectrogram serves as a kind of widely used acoustic feature in MIR tasks such as automatic MGC, music transcription etc. In Section 5.1, we analysed the mel spectrogram of each excerpt by PCA. In Section 5.2, we conducted an automatic classification experiment on the excerpts to discuss the choice of excerpts and compare it with the correlation matrix (Table 3) that can reflect how participants recognise 'Chinese style' in all categories of music chosen in the experiment.

5.1 | Analysis on acoustic features

First, we extracted spectral centroid and spectral bandwidth [19] from the spectrogram of each excerpt (shown in Appendix) to compare their general timbre characteristics. Excerpts played by the violin have higher and more fluctuating spectral centroid than those played by the Erhu, meaning the timbre of violin sounds brighter than that of Erhu. Then, we selected mel spectrogram as the acoustic feature to represent each excerpt and projected it into a low-dimensional space. To see how different durations influence the projection, we selected 2-s, and 5-s randomly[4] of each excerpt to get the mel spectrogram, respectively. PCA [53] was used to reduce the dimension of mel spectrogram sequence to 3 to make the projection results (shown in Figure 5) visible.

In Figure 5a,b, it can be found that excerpts played by the violin and Erhu can be separated apparently in a 3-dimensional space. That means instruments can be easily identified with mel spectrogram in a low-dimensional space for excerpts with merely 2 seconds and longer duration. However, the category of each excerpt cannot be separated in a low-dimensional space. In other words, information related to the music content category (including the Chinese style) failed to be revealed in the dimensionality-reduced mel spectrogram.

5.2 | Experiment results

After initial exploration of acoustic features of each excerpt, we conducted a supplementary experiment of automatic music classification on 8 and 10 types (incorporating Atonal excerpts), respectively, according to the dimension of the correlation matrix (from Section 4) to see if the result can reflect participants' preference shown in the perceptual experiment. The Convolutional Neural Network (CNN) [18, 54, 55] is used as the classifier. The automatic classification results are shown in Table 4.

In the confusion matrix shown in Figure 6, it seems that the classifier rarely mistakes the playing instrument. It achieves better classification performance on samples with longer duration, which contains richer music information. We assume

[4]Randomly select 10 times for 2-s and 5-s

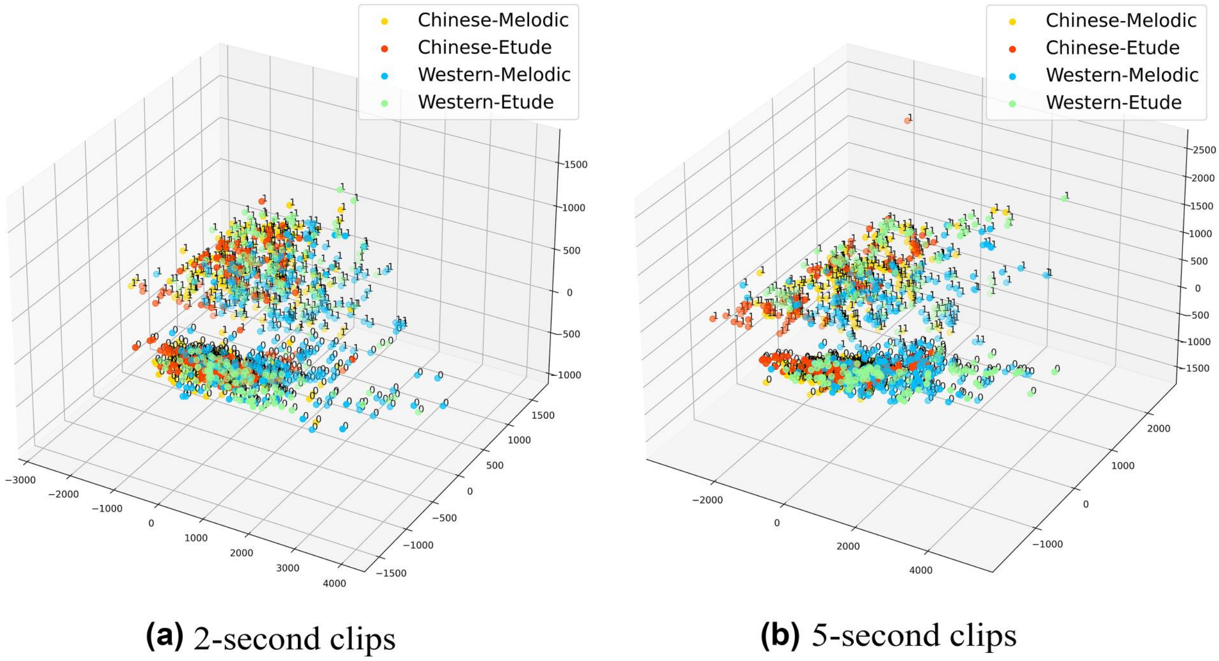


FIGURE 5 Distribution of each excerpt (the number above each point represents the instrument of each excerpt, with 0 as Erhu and 1 as violin; the colour of each point represents its category)

TABLE 4 Results of automatic music classification experiments

Duration	Precision (%)	Recall (%)	F1 (%)
2s (10 types)	76.37	77.05	76.11
5s (10 types)	86.79	85.78	84.80
2s (8 types)	80.58	79.81	79.48
5s (8 types)	94.57	79.81	93.51

that the process of giving ratings of 'Chinese style' among all categories is equal to classification. Then, we compare the people's preference shown in the correlation matrix with this 8-type classification experiment results. Two types with high positive correlation, which participants tend to confuse in the perceptual experiment, are also easy to be confused by the machine. To be specific, as for Chinese style in music, the participants' rating of *Western-Etude* and *Western-Melodic* excerpts reveals high correlation (the red block in Figure 6) and as for classification task, the machine does have a tendency to mistake one certain category as another among the Western excerpts. For example, *Western-Melodic-Violin* excerpts in the correlation matrix are shown to be most positively correlated (0.388) to *Western-Etude-Violin* and negatively correlated (−0.152) to *Chinese-Melodic-Violin*. It can be seen from the confusion matrix that 13.3% of *Western-Melodic-Violin* excerpts are mistaken as *Western-Etude-Violin* and none of them are classified into *Chinese-Melodic-Violin*. The error analysis shows that the process of automatic classification can sometimes reflect people's preference for classification shown in the survey results.

The confusion matrix of 10-type classification experiment are shown in Figure 7. We choose results of epoch 65 and epoch 75 as examples. If we incorporate the *Atonal* category into the classification task, the accuracy is lower than 8-type experiment. This may be due to the imbalance of the data size of each type (shown in Table 5). Specifically in the blue block in Figure 7, about 33.3% of *Atonal* excerpts played by Erhu/violin are mistaken into *Western-Melodic-Erhu/Violin* ones; 16.7% of *Atonal* excerpts played by the Erhu are mistaken into *Atonal* excerpts played by the violin. A possible explanation for this might be that *Atonal* music we chose are composed by Chinese musicians who combined western contemporary composing techniques with traditional pentatonic scale in these excerpts, as mentioned in Section 3.1 [48–50].

6 | CONCLUSION AND FUTURE WORK

In this study, we quantitatively investigated how music content and performed instruments influence the perception of 'Chinese style' in music. We disentangled the two contributing factors by aligning musical scores between Erhu and violin performances. Playing techniques were excluded because they may involve pitch changes. A Wechat Mini programme was

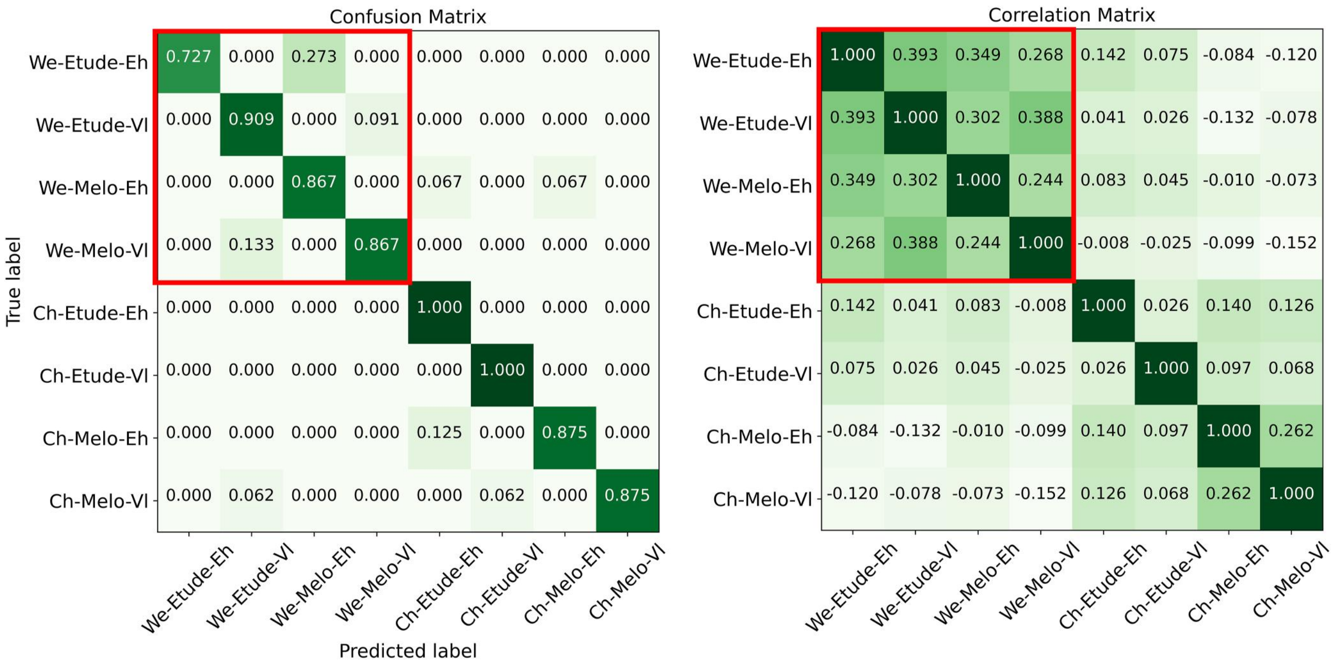


FIGURE 6 The comparison between the confusion matrix of 8-type classification results (5s) and correlation matrix of survey results

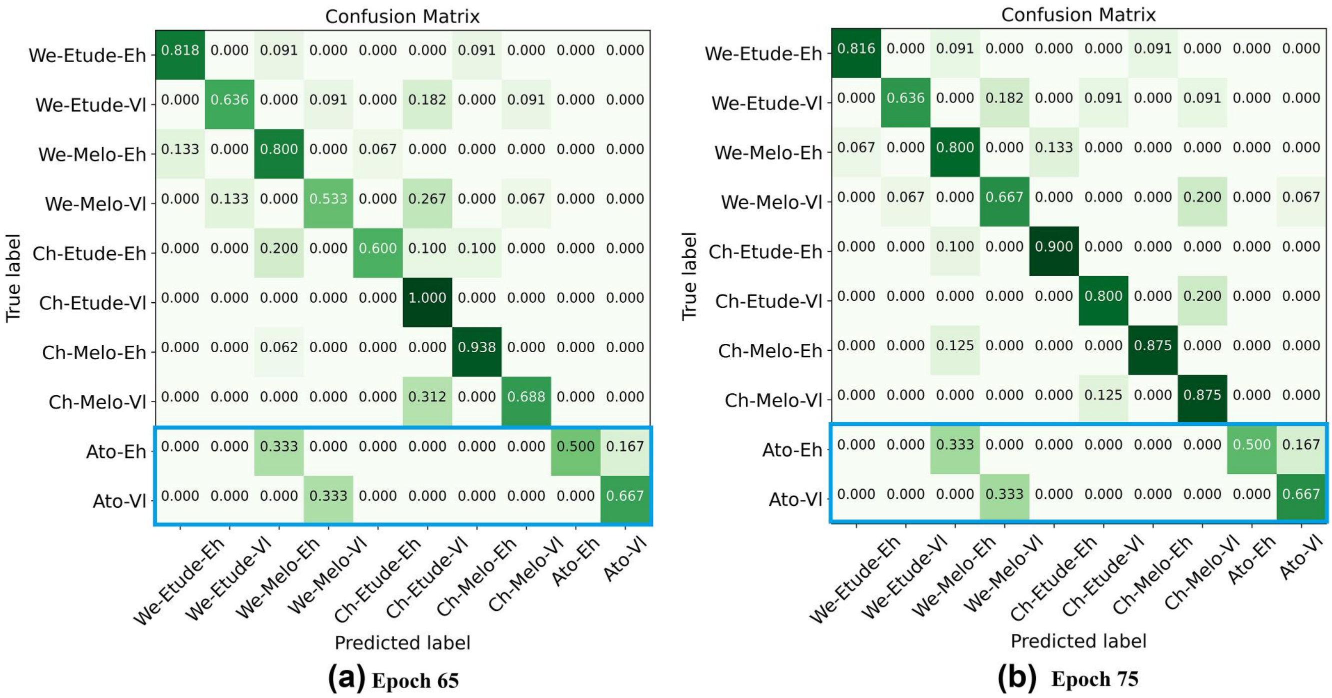


FIGURE 7 Confusion matrix of 10-type classification (5s)

TABLE 5 Statistics of 10 types and their size

Category	Chinese-melodic-Erhu	Chinese-melodic-violin	Western-melodic-Erhu	Western-melodic-violin	Chinese-Etude-Erhu	Chinese-Etude-Violin	Western-Etude-Erhu	Western-Etude-Violin	Atonal (Erhu: 60, violin: 60)
Number	160	160	150	150	100	100	110	110	120

developed to publish the online questionnaire. Six hundred and one responses were collected from participants covering various groups in terms of age, education, and music studies.

Statistical analysis on questionnaire responses reveals that people rely on music content more than instruments when they give ratings of 'Chinese style'. Significant differences were also observed between ratings on excerpts played in the Erhu and the violin, suggesting that people are able to identify 'Chinese style' of the Erhu as a traditional Chinese instrument, although playing techniques were removed. Musicians showed higher sensibility to both music content and instruments and their responses are more concentrated than non-musicians. Ratings on atonal music are similar to western music, which indicates that it's difficult for most people to recognise Chinese traditional tunes hidden in atonal works written by Chinese composers. The questionnaire results reflect people's understandings of Chinese music in the contemporary cultural background. This could be applicable in music recommendation systems and be valuable to musicians in composing Chinese music. In addition, we discussed the choice of involved stimuli, and similarities between computer auditory and human perception on abstract musical characteristics, by conducting an automatic music classification experiment on all excerpts and comparing its results with the perceptual experiment. The findings may motivate the improvement of acoustics features and model structures, making them represent the abstract musical characteristics more comprehensively, which will be helpful for MIR of Chinese music.

The original methodology proposed in this study could be applied in related research. Further interesting experiments can be designed based on this study to study other performance variables that may contribute to the ratings of 'Chinese style'. The playing techniques could be aligned between Erhu and violin performances and quantitative analysis could be carried out. Other instrument families may be considered as well. Besides, data from people with different cultural backgrounds can be collected and compared to explore the cross-cultural musical understandings on 'Chinese style' music. Based on these, we will also study how to disentangle the 'Chinese style' for MIR by improving algorithms, systems, and tools.

ACKNOWLEDGEMENTS

This work was supported by the major project of the National Social Science Fund of China (Grant No. 21ZD19), and this study benefited from support from Xudong Ye, Shanshan Jin, Zengqi Zhang, Yiru Xu, Junran Wang and Haotian Zhou. We also appreciate Dr. Zijin Li for providing valuable feedback during final drafting of the study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Yu Zhang https://orcid.org/0000-0002-0310-3705

REFERENCES

1. Wang, L.: Chinese Music School and its Discourse System Developed. The Art of Music, No 4. 04, 6–17 (2020)
2. Wang, L.: Definition of Chinese Music School. Music Research, No 4. 04, 5–12 (2020)
3. Qin, X.: 中国乐派'的释义与历史定位. Chinese Music. 5, 5–21 (2019)
4. Xia, Y.: 中国乐派'的构成及其传统与方向. Music Research. 4, 13–19 (2020)
5. Lin, X., Bi, M.: 理解'中国乐派'视角转换与范式建构. Music Research. 4, 39–53 (2020)
6. Hu, P.: 中国风格'室内乐界定研究. Modern Music. 4, 19–21 (2019)
7. Yang, H.-L., Saffle, M.: China and the West: Music, Representation and Reception. University of Michigan Press Ann Arbor (2017)
8. Aalst, V.: Chinese Music. Statistical Department of the Inspectorate General of Customs in Shanghai (1884)
9. Tzanetakis, G., et al.: Computational ethnomusicology. J. Interdiscip. Music. Stu. 1(2), 1–24 (2007)
10. Foxworth, N.-A.: Musical Genre Identification and Differentiation of Rock, R&B/Hip-Hop, and Christian Songs Through Harmonic Analysis. Selected honored theses (2017). http://firescholars.seu.edu/honors/74
11. Fabbri, F.: A Theory of Musical Genres: Two Applications, pp. 52–81. Popular Music Perspectives (1981)
12. Van der Merwe, P.: Origins of the Popular Style: The Antecedents of Twentieth-Century Popular Music. Clarendon Press (1989)
13. Dannenberg, R.: Style in music. In: The Structure of Style: Algorithmic Approaches to Understanding Manner and Meaning (2009)
14. Zhu, J., Xue, X., Lu, H.: Musical genre classification by instrumental features. Int. Computer Music Conf. 580–583 (2004)
15. Tzanetakis, G., Essl, G., Cook, P.: Musical genre classification of audio signals. IEEE Trans. Speech Audio Process. 10(5), 293–302 (2002)
16. Baniya, B.-K., Ghimire, D., Lee, J.: Automatic music genre classification using timbral texture and rhythmic content features. In: 17th International Conference on Advanced Communication Technology, pp. 434–443. IEEE, PyeongChang (2015)
17. Ghosal, D., Kolekar, M.H.: Music genre recognition using deep neural networks and transfer learning. In: Proc. Interspeech, 2087–2091 (2018). https://doi.org/10.21437/Interspeech.2018-2045
18. Liu, C., et al.: Bottom-up broadcast neural network for music genre classification. Multimed. Tool. Appl. 80, 7313–7331 (2021)
19. Klapuri, A., Davy, M.: Signal Processing Methods for Music Transcription. Springer, Boston (2006). https://doi.org/10.1007/0-387-32845-9
20. ISMIR. 5th International Conference on Music Information Retrieval (ISMIR 2004) Rhythm Description Contest. (2004). http://mtg.upf.edu/ismir2004/contest/rhythmcontest/
21. Marchand, U., Peeters, G.: The extended Ballroom dataset. In Extended Abstracts for the Late-Breaking Demo Session of the 17th International Society for Music Information Retrieval Conference. New York (2016)
22. Basili, R., Serafini, A., Stellato, A.: Classification of music genre: a machine learning approach. In: Proceedings of 5th International Conference on Music Information Retrieval. Barcelona (2004)
23. McKay, C., Fujinaga, I.: Automatic genre classification using large high-level musical feature sets. In: Proceedings of 5th International Conference on Music Information Retrieval, pp. 525–530. Barcelona (2004)
24. Cataltepe, Z., Yaslan, Y., Sonmez, A.: Music genre classification using MIDI and audio features. EURASIP J. Appl. Signal Process. 1, 150–157 (2007)
25. Ferraro, A., Lemstrom, K.: On large-scale genre classification in symbolically encoded music by automatic identification of repeating patterns. In: Proceedings of 5th International Conference on Digital Libraries for Musicology, pp. 34–37. Association for Computing Machinery New York NY United States, Paris (2018)
26. Oramas, S., Nieto, O., Serra, X.: Multi-label music genre classification from audio, text and images using deep features. In: Proceedings of 185th International Society of Music Information Retrieval Conference, pp. 23–30. Suzhou (2017)
27. Bogdanov, D., Porter, A., Oramas, S.: The AcousticBrainz Genre Dataset: multi-source, multi-level, multi-label, and large-scale. In:

Proceedings of the 20th International Society for Music Information Retrieval Conference, pp. 360–367. Delft (2019)
28. Epure, E.-V., et al.: Modeling the music genre perception across language-bound cultures. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 4765–4779. Association for Computational Linguistics, Punta Cana, Dominican Republic (2020)
29. Lena, J.-C.: Banding Together: How Communities Create Genres in Popular Music. Princeton University Press, Princeton (2012)
30. Ferwerda, B., Tkalcic, M., Schedl, M.: Personality traits and music genres: what do people prefer to listen to. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, pp. 285–288. Bratislava (2017)
31. Sordo, M., et al.: The quest for musical genres: do the experts and the wisdom of crowds agree? In: Proceedings of 9th International Society for Music Information Retrieval, pp. 255–260. (2008)
32. Pati, A., Gururani, S., Lerch, A.: dMelodies: a music dataset for disentanglement learning. In: Proceedings of the 21st International Society for Music Information Retrieval Conference. Montréal, Canada (2020)
33. Luo, Y.-J., et al.: Unsupervised disentanglement of pitch and timbre for isolated musical instrument sounds. In: Proceedings of the 21st International Society for Music Information Retrieval Conference. Montréal, Canada (2020)
34. Lee, J., et al.: Metric learning vs classification for disentangled music representation learning. In: Proceedings of the 21st International Society for Music Information Retrieval Conference. Montréal, Canada (2020)
35. Tian, M., et al.: Towards the representation of Chinese traditional music: a state of the art review of music metadata standards. In: International Conference on Dublin Core and Metadata Applications, pp. 71–81. Open Journal Systems, Portugal (2013)
36. Repetto, R.-C., Serra, X.: A collection of music scores for corpus based Jingju singing research. In: Proceedings of the 18th International Society for Music Information Retrieval Conference, pp. 50–52. Suzhou, China (2017)
37. Gong, R., Repetto, R.-C., Serra, X.: Creating an A Cappella singing audio dataset for automatic Jingju singing evaluation research. In: Proceedings of the 4th International Workshop on Digital Libraries for Musicology, pp. 37–40. Association for Computing Machinery New York NY United States, Shanghai (2017)
38. Repetto, R.-C., Serra, X.: Quantitative analysis of the relationship between linguistic tones and melody in Jingju using music scores. In: Proceedings of the 4th International Workshop on Digital Libraries for Musicology, pp. 41–44. Association for Computing Machinery New York NY United States, Shanghai (2017)
39. Zhang, S., Repetto, R.-C., Serra, X.: Understanding the expressive functions of jingju metrical patterns through lyrics text mining. In: Proceedings of 18th International Society for Music Information Retrieval Conference, pp. 397–403. Suzhou, China (2017)
40. Poulin-Charronnat, B., et al.: Effects of a change in instrumentation on the recognition of musical materials. Music Perception. 22(2), 239–263 (2004)
41. McAdams, S., et al.: Perception of musical similarity among contemporary thematic materials in two instrumentations. Music Perception. 22(2), 207–237 (2004)
42. Wolpert, R.-S.: Recognition of melody, harmonic accompaniment, and instrumentation: musicians vs. Nonmusicians. Music Perception. 8(1), 95–105 (1990)
43. Novello, A., McKinney, M.M.F., Kohlrausch, A.: Perceptual evaluation of intersong similarity in Western popular music. J. N. Music Res. 40(1), 1–26 (2011)
44. Halpern, A.R., Daniel, M.: Effects of timbre and tempo change on memory for music. Q. J. Exp. Psychol. 61(9), 1371–1384 (2008)
45. Schellenberg, E.G., Stalinski, S.M., Marks, B.M.: Memory for surface features of unfamiliar melodies: independent effects of changes in pitch and tempo. Psychol. Res. 78(1), 84–95 (2014)
46. Demorest, S.M., et al.: Lost in translation: an enculturation effect in music memory performance. Music Perception. 25(3), 213–223 (2008)
47. Inesta, J.-M., Ponce de Leó n, P.-J., Heredia, J.-L.: A ground-truth experiment on melody genre recognition in absence of timbre. In: Proceedings of the 10th International Conference on Music Perception and Cognition, pp. 758–761. Sapporo (2008)
48. Zheng, Y.: 桑桐的《夜景》试释. Art of Music (02), 54–59 (1983)
49. Peng, M.: 浅析我国第一首无调性作品——桑桐的小提琴独奏曲《夜景》. 大众文艺, (21), 66–67 (2009)
50. Wang, J.: Origin Out of the People, Basis on the Traditions: Notes on the Composition of the Second Erhu Rhapsody, People's Music 09, pp. 2–5 +63. (2003)
51. Philip, H.R.: Critical values for Spearman's rank order correlation. J. Educ. Behav. Statist. 14(3), 245–253 (1989)
52. Stevens, S.-S., Volkmann, J.: A scale for the measurement of a psychological magnitude: loudness. Psychol. Rev. 43(5), 405–416 (1936)
53. Pearson, K.: On lines and planes of closest fit to systems of points in space. Phil. Mag. 2(11), 559–572 (1901)
54. LeCun Y., et al.: Object recognition with gradient-based learning. In: Shape, Contour and Grouping in Computer Vision. Lecture Notes in Computer Science, vol 1681. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-46805-6_19
55. Lecun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE. 86(11), 2278–2324 (1998)

How to cite this article: Zhang, Y., Zhou, Z., Sun, M.: Influence of musical elements on the perception of 'Chinese style' in music. Cogn. Comput. Syst. 4(2), 147–164 (2022). https://doi.org/10.1049/ccs2.12036

APPENDIX

I Online wechat Mini-program development

This section describes the specific process and technical implementation details of the questionnaire survey

(i) Off-site (online) experiment

On-site testing are common approaches of music perception experiments that require high-quality sound effects. In this experiment, we choose off-site experiments due to the relatively low requirements of the playback equipment. Headphones are recommended and cell phone speakers should be fine if participants are in quiet environment. The off-site (online) experiment has the following advantages.

(1) Allow subjects to participate in experiments with electronic devices (phones, tablets, computers etc.) anywhere and anytime, as long as with the access to the Internet.
(2) Easy to disseminate among people with various backgrounds.
(3) Aims to collect greater number of responses than on-site experiment

Considering the off-site technical possibilities, we chose WeChat Mini programme as the online experiment platform with the following benefits. (1) In Q1 2021, WeChat had 1.24 billion active users, the only app in China to have over one billion active users.[5] (2) Compared with web-based development, WeChat mini programs have a shorter development cycle.

On the other hand, the disadvantage of off-site experiment is obvious: the testing flow of the process cannot be

[5]https://www.businessofapps.com/data/wechat-statistics/

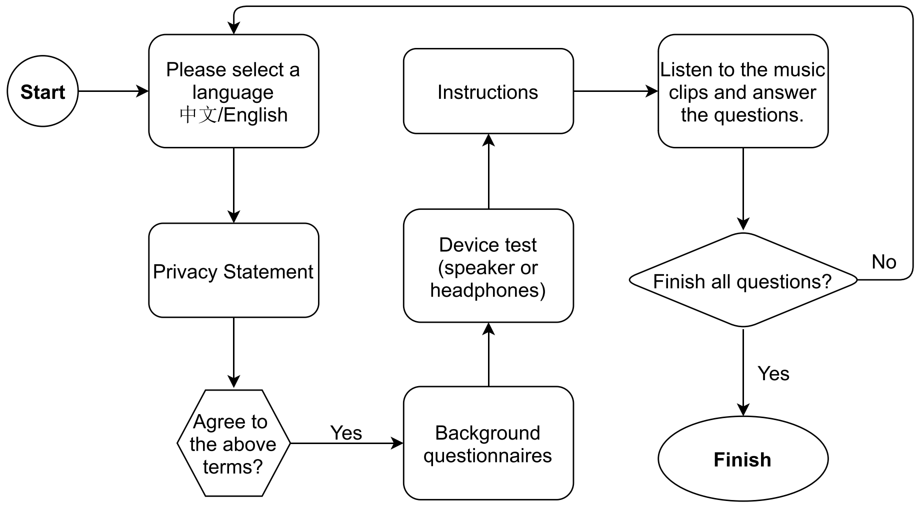


FIGURE A1 Flow chart of the online questionnaire

guaranteed like the on-site experiment. For example, participants may pause, stop or fast-forward when they listen to the music, which may include invalid responses and thus cause reliability loss of the survey results. A series of measures were designed and carried out to solve the above potential problems and thus ensure the valid responses (see Section 3).

(ii) The experiment flow.

To start with, the users need to select a language and agree with the privacy statement. Before the listening test, users' background information regarding music education was collected. Then they read the instructions and test the speaker/headphone normal function. After finishing all listening tests, they should quit the programme. The overall test lasts 5–7 min. Only complete questionnaires responses were used as the valid results. The flow chart of the online questionnaire is shown in Figure A1.

(iii) Implementation details to ensure valid responses.

To ensure the consistency of listening procedures and avoid users' toggling with the player buttons, all music excerpts were automatically played without any buttons displayed. They were not able to answer the questions until the playing finished (12 s on average). After clicking on the buttons to answer the questions, they immediately jump to the next listening item, having no chances to reconsider their choices. This ensured that they answered based on the first intuition. Some common sense questions were interspersed every 3–4 listening items, for improving participants' concentration.

II List of stimuli

Excerpt	Composer	Metre	Tonality/Mode	Duration	BPM	Year of composition	Region
Chinese-Melodic							
Bumper Harvest (丰收) part1	Wang Yi	2/4	D Gong(宫) Qing Yue Mode (清乐)	13s	150	1953	JiangSu
Bumper Harvest (丰收) part2	Wang Yi	2/4	D Gong(宫) Qing Yue Mode (清乐)	20s	60	1954	JiangSu
Bumper Harvest (丰收) part3	Wang Yi	2/4	D Gong(宫) Qing Yue Mode (清乐)	17s	60	1955	JiangSu
Moon Night (月夜) part1	Liu Tianhua	4/4	D Gong(宫) hexatonic (Bian Gong)	20s	40	1918	China Southern
Moon Night (月夜)part2	Liu Tianhua	4/4	D Gong(宫) hexatonic (变宫)	20s	60	1918	China Southern
Autumn Moon over the Han Palace (汉宫秋月) part1	Ancient Tune Score by Liu Tianhua	2/4	D Gong(宫) Ya Yue Mode (雅乐)	20s	60	Yuan Dynasty	China
Autumn Moon over the Han Palace (汉宫秋月) part2	Ancient Tune Score by Liu Tianhua	2/4	D Gong(宫) Ya Yue Mode (雅乐)	14s	60	Yuan Dynasty	China
Autumn Moon over the Han Palace (汉宫秋月) part3	Ancient Tune Score by Liu Tianhua	2/4	D Gong(宫) Ya Yue Mode (雅乐)	23s	60	Yuan Dynasty	China

APPENDIX (Continued)

Excerpt	Composer	Metre	Tonality/Mode	Duration	BPM	Year of composition	Region
Mihu Melody(迷胡调) part1	Lu Rirong	2/4	D Gong(宫) Qing Yue Mode (清乐)	19s	100	1958	Shanxi
Mihu Melody(迷胡调) part2	Lu Rirong	2/4	D Gong(宫) Qing Yue Mode (清乐)	20s	100	1958	Shanxi
Recitation of Leisure (闲居吟) part1	Liu Tianhua	4/4	D Gong(宫) Qing Yue Mode (清乐)	21s	40	1928	China
Recitation of Leisure (闲居吟) part2	Liu Tianhua	4/4	D Gong(宫) Qing Yue Mode (清乐)	17s	60	1928	China
Recitation of Leisure (闲居吟) part3	Liu Tianhua	4/4	D Gong(宫) Qing Yue Mode (清乐)	20s	60	1928	China
With Happy Songs Sung (幸福的歌儿唱不完) part1	Su Anguo	2/4	D Yu(羽) Ya Yue (雅乐)	10s	60	20th century	Shandong China
With Happy Songs Sung (幸福的歌儿唱不完) part2	Su Anguo	2/4	D Yu(羽) Ya Yue (雅乐)	18s	60	20th century	Shandong China
With Happy Songs Sung (幸福的歌儿唱不完) part3	Su Anguo	3/4	D Yu(羽) Ya Yue (雅乐)	17s	60	20th century	Shandong China
Chinese-Etude							
Erhu etude no. 8	Liu Tianhua	2/4	D Gong(宫) hexatonic (变宫)	17s	80	1920s	China
Erhu etude no. 10	Liu Tianhua	2/4	D Gong(宫) Qing Yue Mode (清乐)	18s	80	1920s	China
Erhu etude no. 11	Liu Tianhua	2/4	D Gong(宫) Qing Yue Mode (清乐)	18s	80	1920s	China
Erhu etude no. 14	Liu Tianhua	2/4	D Gong(宫) Qing Yue Mode (清乐)	16s	80	1920s	China
Erhu etude no. 16	Liu Tianhua	2/4	D Gong(宫) hexatonic (变宫)	16s	80	1920s	China
Erhu etude no. 25	Liu Tianhua	2/4	D Gong(宫) Qing Yue Mode (清乐)	12s	60	1920s	China
Erhu etude no. 26	Liu Tianhua	2/4	D Gong(宫) Qing Yue Mode (清乐)	14s	80	1920s	China
Erhu etude no. 31	Liu Tianhua	2/4	D Gong(宫) Qing Yue Mode (清乐)	18s	80	1920s	China
Erhu etude no. 42	Liu Tianhua	2/4	D Gong(宫) Qing Yue Mode (清乐)	21s	80	1920s	China
Erhu etude no. 44	Liu Tianhua	2/4	D Gong(宫) Qing Yue Mode (清乐)	17s	80	1920s	China
Western-Melodic							
Op. 89 Aires Variés No. 6	Charles Dancla	4/4	D major	24s	80	1858	France
Op. 89 Aires Variés No. 6	Charles Dancla	4/4	D major	15s	60	1858	France
Op. 89 Aires Variés No. 6	Charles Dancla	4/4	D major	17s	60	1858	France
Concerto No. 5, 1st Mvt.part1	Friedrich Seitz	4/4	D major	15s	60	1909	Germany
Concerto No. 5, 1st Mvt.part2	Friedrich Seitz	4/4	D major	17s	60	1909	Germany
Concerto No. 5, 1st Mvt.part3	Friedrich Seitz	4/4	D major	15s	60	1909	Germany
Mozart B flat Major part 1	Wolfgang Amadeus Mozart	4/4	B♭ major	19s	60	18th century	Austria
Mozart B flat Major part 2	Wolfgang Amadeus Mozart	4/4	B♭ major	24s	60	18th century	Austria
Sonate 5 part 1	Wolfgang Amadeus Mozart	4/4	E harmonic minor	16s	60	18th century	Austria
Sonate 5 part 2	Wolfgang Amadeus Mozart	4/4	E harmonic minor	16s	40	18th century	Austria
Sonate 5 part 3	Wolfgang Amadeus Mozart	4/4	E harmonic minor	23s	60	18th century	Austria
Sonate 6 part 1	Wolfgang Amadeus Mozart	4/4	B♭ major	16s	60	18th century	Austria

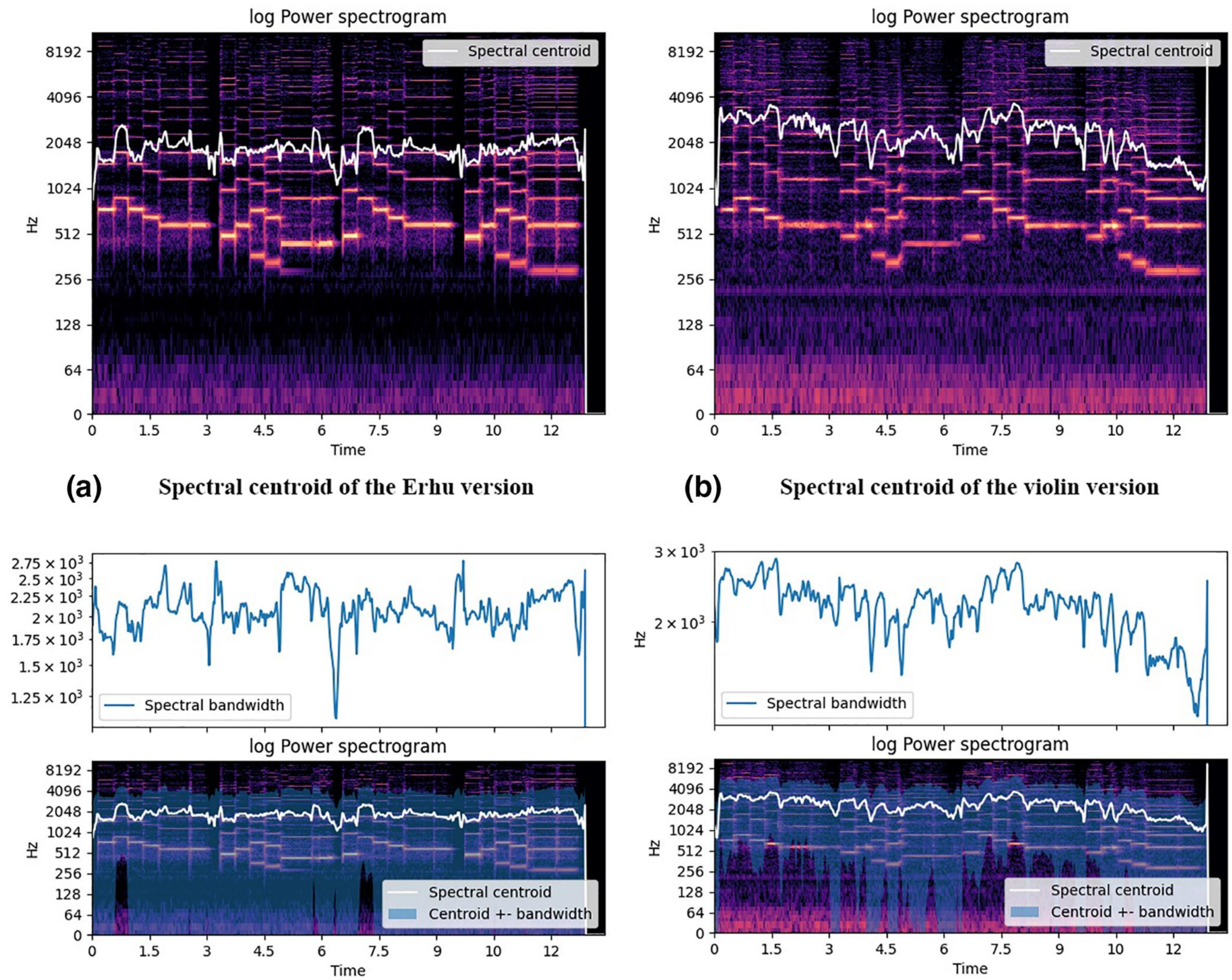
(Continues)

APPENDIX (Continued)

Excerpt	Composer	Metre	Tonality/Mode	Duration	BPM	Year of composition	Region
Sonate 6 part 2	Wolfgang Amadeus Mozart	4/4	B♭ major scale	9s	60	18th century	Austria
Sonate 6 part 3	Wolfgang Amadeus Mozart	4/4	B♭ major scale	24s	60	18th century	Austria
Sonate 6 part 4	Wolfgang Amadeus Mozart	4/4	B♭ major scale	11s	60	18th century	Austria
Western-Etude							
Kayser Op.20.4 part 1	Heinrich Ernst Kayser	4/4	C major	12s	80	17th century	Germany
Kayser Op.20.4 part 2	Heinrich Ernst Kayser	4/4	C major	12s	80	17th century	Germany
Kayser Op.20.13 part 1	Heinrich Ernst Kayser	3/4	G major	14s	80	17th century	Germany
Kayser Op.20.13 part 2	Heinrich Ernst Kayser	3/4	G major	12s	80	17th century	Germany
Kayser Op.20.13 part 3	Heinrich Ernst Kayser	3/4	G major	11s	80	17th century	Germany
Kayser Op.20.22	Heinrich Ernst Kayser	4/4	E major	11s	80	17th century	Germany
Kayser Op.20.23	Heinrich Ernst Kayser	2/4	F harmonic minor	21s	80	17th century	Germany
Perception Motion part 1	Ottokar Novacek	4/4	C major	16s	60	17th century	Hungary
Perception Motion part 2	Ottokar Novacek	4/4	C major	17s	60	17th century	Hungary
Perception Motion part 3	Ottokar Novacek	4/4	C major	17s	60	17th century	Hungary
Perception Motion part 4	Ottokar Novacek	4/4	C major	17s	60	17th century	Hungary
Atonal							
Erhu Rhapsody No. 2 part 1	Wang Jianmin	2/4	Yu mode extended by artificial scales (Hunan huagu Opera)[50]	24s	40	2001	Hunan, China
Erhu Rhapsody No. 2 part 2	Wang Jianmin	4/4	Yu mode extended by artificial scales (Hunan huagu Opera)[50]	23s	40	2001	Hunan, China
Erhu Rhapsody No. 2 part 3	Wang Jianmin	4/4	Yu mode extended by artificial scales (Hunan huagu Opera)[50]	20s	40	2001	Hunan, China
Night Scene part 1	Sang Tong	multi	Serial	20s	60	1947	China
Night Scene part 2	Sang Tong	multi	Serial	21s	60	1947	China
Night Scene part 3	Sang Tong	multi	Serial	18s	60	1947	China

III Examples of acoustics features

Excerpt: Bumper Harvest (丰收) composed by Wang Yi (Figure A2).



(a) Spectral centroid of the Erhu version

(b) Spectral centroid of the violin version

(c) Spectral bandwidth of the Erhu version

(d) Spectral bandwidth of the violin version

FIGURE A2 Spectral centroid and spectral bandwidth of the excerpt from *Fengshou*-Yi Wang

IV Correlation matrix of atonal excerpts

TABLE A1 Correlation matrix of atonal excerpts

Category	Western-Etude-Erhu(1)	Western-Etude-Violin(2)	Western-Melodic-Erhu(3)	Western-Melodic-Violin(4)	Chinese-Etude-Erhu(5)	Chinese-Etude-Violin(6)	Chinese-Melodic-Erhu(7)	Chinese-Melodic-Violin(8)	Atonal-Erhu(9)	Atonal-Violin(10)
Western-Etude-Erhu(1)	1	**0.393**	**0.349**	**0.268**	**0.142**	0.075	**-0.084**	**-0.12**	-0.025	0.095
Western-Etude-Violin(2)	**0.393**	1	**0.302**	**0.388**	0.041	0.026	**-0.132**	-0.078	0.113	0.039
Western-Melodic-Erhu(3)	**0.349**	**0.302**	1	**0.244**	**0.083**	0.045	-0.01	-0.073	0.092	-0.016
Western-Melodic-Violin(4)	**0.268**	**0.388**	**0.244**	1	-0.008	-0.025	**-0.099**	**-0.152**	0.056	0.147
Chinese-Etude-Erhu(5)	**0.142**	0.041	**0.083**	-0.008	1	0.026	**0.14**	**0.124**	0.089	0.184
Chinese-Etude-Violin(6)	0.075	0.026	0.045	-0.025	0.026	1	**0.095**	0.068	-0.037	-0.007
Chinese-Melodic-Erhu(7)	**-0.084**	**-0.132**	-0.01	**-0.099**	**0.14**	**0.095**	1	**0.262**	-0.017	0.018
Chinese-Melodic-Violin(8)	**-0.12**	-0.078	-0.073	**-0.152**	**0.124**	0.068	**0.262**	1	0.063	-0.059
Atonal-Erhu(9)	-0.025	0.113	0.092	0.056	0.089	-0.037	-0.017	0.063	1	0.074
Atonal-Violin(10)	0.095	0.039	-0.016	0.147	0.184	-0.007	0.018	-0.059	0.074	1

Note: No significant correlations were observed between atonal categories and others($p > 0.05$). N(Atonal-Erhu) = 378, N(*Atonal-violin*) = 223.