JOINT ANALYSIS OF MODE AND PLAYING TECHNIQUE IN GUQIN PERFORMANCE WITH MACHINE LEARNING

Yu-Fen Huang[1]   **Jeng-I Liang**[2]   **I-Chieh Wei**[1]   **Li Su**[1]

[1]Institute of Information Science, Academia Sinica, Taiwan
[2]Department of Traditional Music, Taipei National University of the Arts, Taiwan

`{yfhuang, sma1033, lisu}@iis.sinica.edu.tw, liangjengi@gmail.com`

ABSTRACT

Music is hierarchically structured, in which the global attributes (e.g., the determined tonal structure, musical form) dominate the distribution of local elements (e.g., pitch, playing technique arrangement). Existing methods for instrumental playing technique detection mostly focus on the local features extracted from audio. However, we argue that structural information is critical for both global and local tasks, particularly considering the characteristics of guqin music. Incorporating mode and playing technique analysis, this study demonstrates that the structural relationship between notes is crucial for detecting mode, and such information also provides extra guidance for the playing technique detection in local-level. In this study, a new dataset is compiled for guqin performance analysis. The mode detection is achieved by pattern matching, and the predicted results are conjoined with audio features to be inputted into a neural network for playing technique detection. Advanced techniques are developed to optimize the extracted pitch contour from the audio. It is manifest in the results that the global and local features are inter-connected in guqin music. Our analysis identifies key components affecting the recognition of mode and playing technique, and challenging cases resulting from unique properties of guqin audio signal are discussed for further research.

1. INTRODUCTION

The guqin (古琴) is a plucked seven-string Chinese musical instrument existing for over 3,000 years, and has been selected as UNESCO World Cultural Heritage.[1] In guqin performance, it is an intrinsic convention for musicians to implement diverse playing techniques, in order to communicate their interpretations of the performed music. In the theory of guqin performance, such configuration of playing technique in local-level is considered to be connected and reflect the higher-level, hierarchical tonal structure in

---

[1] https://ich.unesco.org/en/RL/guqin-and-its-music-00061

pentatonic modes [1]. However, existing research does not appear to provide sufficient empirical evidence to support the theory. In particular, in the research area of Music Information Retrieval (MIR), key/mode detection has been regarded as the recognition of the overall, global construction of music piece, whereas playing technique detection is usually taken as a classification task relying on local features extracted from the audio. This study aims to solve this discontinuity, and argue that the global tonal structure and the local configuration of playing technique are interconnected in guqin performance. This work takes MIR and machine learning frameworks as the means for empirical music analysis, and contribute to several aspects including:

1) to design and compile a new dataset, GQ39, featured by representative historical guqin recordings and note-by-note annotations;
2) to demonstrate the importance of tonal structural information, and to identify the crucial components contributing to both the mode detection and playing technique detection;
3) to bridge the knowledge gap between the theory and empirical observations, as well as to highlight the connection between the high-level structure and local elements in music.

In the subsequent section, related research regarding the tonal structure investigation, playing technique classification, and guqin performance theories will be reviewed. The theoretical basis and the procedure to construct the GQ39 dataset will be described in Section 3. The mode detection is performed on the dataset, and the results are analyzed in Section 4. The playing technique is further investigated using a neural network, and decisive components playing important roles in the task are discussed in Section 5, followed by the concluding remarks in Section 6.

2. RELATED WORK

For the high-level tonal structure in music, key detection has been one of the core issues to be explored in Western tonal music. Methods based on chord progression rules are applied to detect key modulation in audio [2]. Neural networks are utilized for key classification, and it has been found that the global harmonic structure in the whole piece plays an important role to identify local keys of short segments [3]. In addition to the connection between the local and global tonal structures, it has also been proven that

different music styles are more easily to be classified by informing the global key and key-related pitch classes [4]. In Indian art music, the usage of raga achieves the structural coherence in music, and different types of raga can be identified according to their pitch distribution [5]. The implicit patterns of melodic and timing features in raga has also been explored [6]. In Arab-Andalusian music, its centonization is analyzed using a high-order n-gram model [7], and different music patterns, nawba, can be classified using template matching [8]. In Georgian music, its unique tuning system has been examined through melodic and harmonic aspects, and the structural relationship between pitch intervals has been found [9].

In previous studies of playing technique classification, diverse features are extracted from audio according to the traits of individual instruments. In particular, strings appear to be high-profile instruments in this research area. For electric guitar, note-level timbral and pitch features are considered to be major components for identifying different types of playing techniques [10], whereas for guitar, each playing technique possesses distinguishable cepstral and phase features [11]. In a study to classify plucking styles of electric bass guitar, intra-note attributes corresponding to the attack and decay parts of the notes are further examined using spectral and statistical descriptors [12]. In violin performance, note-level features including dynamics, vibrato rate and vibrato extent are proven to be connected with score-informed expressive schemes [13], and idiosyncratic performing styles of individual violinists can be characterized by their articulation, energy, and vibrato attributes [14].

In guqin performance, the fingering and playing techniques for strings pressing in the left hand have been regarded as a primary component to affect the expression and aesthetic perception of performance, especially since the late Ming dynasty (around the late 17th century) [15–17]. Furthermore, the left-hand techniques are not merely local ornaments attached to single musical events, and should be contemplated within the global context of complete music composition. As stated in conventional guqin performance theory, the hierarchical tonal structure in the mode can vastly affect the selection of playing technique in the practice of guqin performance [1, 18, 19]. The connection between the global tonal structure and local technique elements is manifest in guqin music theories, and such association between the mode and playing technique is also a prevailing, shared character in many Asian music cultures [20,21]. However, only few efforts have been devoted to explore empirical evidence regarding how the global and local aspects conjoin with one another in the actual performance practice [22, 23]. This study therefore aims to bridge the knowledge gap between the guqin music theories and the empirical observation, and explore the connection between the mode structure and the playing technique implement using statistics and machine learning.

3. GQ39 DATASET

Guqin music is constructed from diverse modes of pentatonic scale, and the performance carries distinctive ex-

| Note degree | $\hat{1}$ | $\hat{2}$ | $\hat{3}$ | $\hat{5}$ | $\hat{6}$ |
| Note name | 宫(Gong) | 商(Shang) | 角(Jue) | 徵(Zhi) | 羽(Yu) |

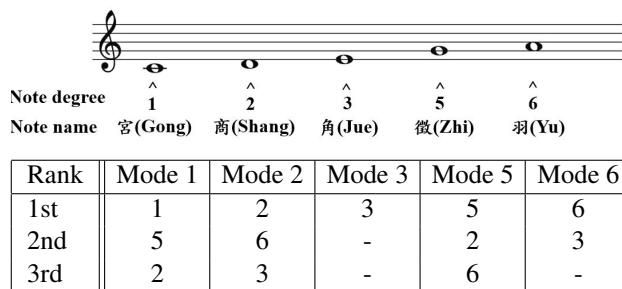| Rank	Mode 1	Mode 2	Mode 3	Mode 5	Mode 6
1st	1	2	3	5	6
2nd	5	6	-	2	3
3rd	2	3	-	6	-

Table 1. The hierarchical structure in pentatonic scale (the top row): the 3 importance level for note degrees (the bottom row) in 5 modes (the middle row).

pressions in playing techniques. We therefore design and compile a new dataset for music performance analysis according to such properties.

3.1 The guqin performance and pentatonic scale

Guqin performers pluck strings by their right hands and press strings using left hands for performing. Guqin performance is characterized by its flexibility, which render plenty of freedom for musicians to choose a wide selection of playing techniques to perform the same piece of music. In particular, the usage of vibrato and portamento is central among all playing techniques to carry the representative traits of individual music pieces and musicians.

The pentatonic scale (see the top row of Table 1 as an example) forms the main construction of guqin music. While other altered chromatic tones and microtones can be included, the pentatonic scale retain the central position in guqin compositions. The pentatonic scale can be transformed into 5 different modes, each of which has a different order of notes according to its initial degree: mode 1 (mode Gong (宫); degree 1, 2, 3, 5, 6), mode 2 (mode Shang (商); degree 2, 3, 5, 6, 1), mode 3 (mode Jue (角); degree 3, 5, 6, 1, 2), mode 5 (mode Zhi (徵); degree 5, 6, 1, 2, 3), and mode 6 (mode Yu (羽); degree 6, 1, 2, 3, 5). Each mode has its own hierarchical structure, which is achieved by assigning different importance levels to note degrees based on the Circle of Fifths. In each mode, at most 3 notes can be considered as prominent component (equivalent concept to the Tonic, Dominant, and Subdominant in Western major and minor scales, see the lower two rows in Table 1) [1, 19]. It should be noted that the pentatonic scale only denotes the relative intervals between notes, but not the absolute pitches, which means that the degree 1 in the scale can be assigned to any pitch in 12 semitones.

3.2 Data collection and labelling

The GQ39 dataset consists of 39 audio recordings of prevalent guqin solo compositions and corresponding event-by-event annotations. The audio data are extracted from a massive collection of guqin historical recording, with the recording years ranging from 1960 to 1990 [24]. 39 excerpts played by five different guqin performers are se-

lected according to a professional guqin musician's opinion (with the performing experience of 18 years). The 39 selected audio excerpts cover representative composing and performing styles and contain diverse playing techniques. The GQ39 dataset only includes guqin music in mode 1, 2, and 5, on account of the facts that: 1) the 2 modes with less than 3 important notes (mode 3 and mode 6) are considered as having less stable structure; 2) mode 3 is only occasionally applied to guqin composition; 3) guqin music composed by mode 6 is relatively complicated, which is not applicable for this exploratory study. The collected audio recordings are roughly 2,000 seconds long in total.

Considering the typical feature of guqin playing, we define an event as a plucking movement in the right hand, which may correspond to either one single note, or a series of note pitch variations and sliding movements in the left hand (ranging up to 5 semitones). Each event is then annotated with 13 types of label by professional guqin musicians, including 2 music-level features (tuning, mode), and 11 event-level features regarding the event context (note degree, pitch range), playing techniques in the right hand for plucking (plucked string, plucking finger, plucking technique type), playing techniques in the left hand for string pressing (pressed position, pressing finger, technique type for pitch variation, technique type for timbre variation), and performing matters (event onset, event duration). As the result, the GQ39 dataset comprises 2,303 annotated events in total (mean # of event = 59.1, SD = 29 per excerpt). The annotations are available on the website, together with details regarding the annotation procedure and dataset descriptions. [2]

4. MODE DETECTION AND ANALYSIS

In this section, we examine mode, the global tonal structure in guqin performance. Chromas are derived from audio recordings. Two types of template representing the tonal structure are designed for pattern matching. The statistical analysis reveals that the inherent characters of mode construction are reflected in the performance. And the results of mode detection indicate that the hierarchical configuration is a crucial element to identify different modes.

4.1 Data representation and mode matching

In mode detection, three types of data are obtained from audio data: the chroma representation of constant-Q transform (CQT), pitch salience function, and pitch contour. For the first type of data, the chromagrams of audio recordings are derived using CQT [25]. For the second and the third types of data, the pitch estimation network, Crepe, is applied to estimate the pitch salience function and the pitch contour [26]. For all types of data, we obtain 60 chromas instead of 12 chromas per octave for higher pitch resolution, considering the facts that: 1) the tuning in guqin is not equal temperament, and 2) the dataset contains large amounts of protamento and transitions between semitones.
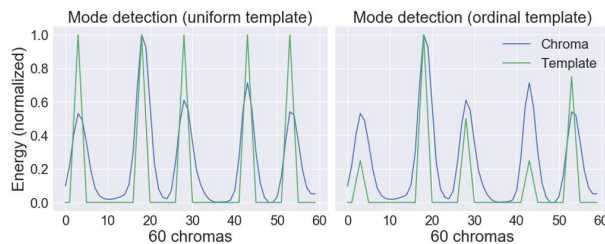
Figure 1. The mode detection with 2 template types: the uniform template (left) and the ordinal template (right). This instance shows the mode 5 template, with degree 5 being shifted to the position of 18 in x-axis.

For each recording, the energy of each chroma is accumulated and is normalized to the range between 0 and 1.

We design two types of mode template, i.e., the uniform template and the ordinal template (see Figure 1 as examples), to be matched up with the chroma representation from audio. For each mode, the *uniform mode template* is constructed by denoting the positions of note degrees as 1 and otherwise 0 in 12 semitones (e.g., mode 1 = (1,0,1,0,1,0,0,1,0,1,0,0)). In addition to the degree positions, the *ordinal mode template* includes the information regarding the importance level of degrees. The degree positions are represented as 4, 3, 2 individually following the order of importance ranking in Table 1, and the two trivial degrees are marked as 1, otherwise 0 in 12 semitones (e.g., mode 1 = (4,0,2,0,1,0,0,3,0,1,0,0)). The two 12-D templates are extended to 60-D for the subsequent matching with 60-chroma representation (each chroma unit represents 20 cents). Complying the treatment procedure for audio chroma representation, all mode templates are normalized to the range between 0 and 1.

The chroma representation of each music excerpt is then matched with mode templates. The built mode templates are shifted to the position of 60 chromas in turn, and are then compared with the audio chroma representation using Pearson correlation. This procedure results in 360 matching pairs for per music excerpt (2 template types (uniform/ordinal) x 3 modes (mode1/2/5) x 60 chroma shifts). The mode is determined by the matching with the highest Pearson correlation coefficient.

4.2 Results and discussion

The statistics of the annotated note degrees reveal the structural configuration of modes, and such observation provides insights regarding how the music theory is practiced in actual performances. As stated in conventional music theories, mode 1 holds the most solid construction among all modes [18]. As can be observed in Figure 2, the note degrees with higher importance ranking usually occur more frequently in the composition (such as degree 1, 5, 2 in mode 1; degree 2, 6, 3 in mode 2; degree 5, 2, 6 in mode 5). Furthermore, in mode 1, the prominent roles of the three principal notes are settled by obvious differences of occurrence ratios, compared to two other trivial degrees. In the mode detection task, we can take a step
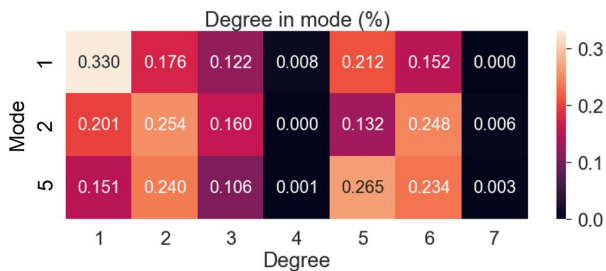
Figure 2. The statistics of the annotated degrees (x-axis) in different modes (y-axis) (in %).

Type	CQT		Salience		Contour	
Result	uni	ord	uni	ord	uni	ord
Correct #	14	**34**	19	29	17	30
Fifth #	4	**4**	5	8	6	8
Miss #	21	**1**	15	2	16	1
Accuracy	0.41	**0.92**	0.55	0.84	0.51	0.87

Table 2. The results of mode detection task (3 data types (CQT/ salience function/ pitch contour) x 2 template types (uniform/ ordinal)).

further to evaluate that how the structural configuration of mode affects the task outcome.

For the evaluation of mode detection, we follow the weighted accuracy used in the audio key detection campaign in MIREX.[3] The instance is marked as 1 point when the prediction is the same mode as the ground truth, whereas 0.5 points is marked when the predicted mode is in perfect fifth above the ground truth (i.e. mode 1 is predicted as mode 5). The results are shown in Table 2. It is evident that all the tasks with ordinal mode template outperform their counterparts applying uniform template. In fact, the *t*-test for the prediction accuracy using uniform/ ordinal template yields a two-tailed *p*-value of 0.0013, which indicates a significant difference between the tasks adopting/ without the structural configuration of mode. To incorporate the statistics of annotated notes with the results of mode detection, above analyses lead to the finding that the information regarding the hierarchical configuration in mode contributes vastly in the mode detection.

5. PLAYING TECHNIQUE DETECTION AND ANALYSIS

In this section, we investigate the classification of left-hand playing technique in guqin performance, and also examine the interaction between global and local features during the classification. Frame-level together with mid-level and high-level features are obtained from audio data and annotations. Dynamic programming is applied to optimize the extracted pitch contour. Six types of playing techniques are then classified using a neural network. The statistical analysis indicates that the hierarchical tonal configuration is embedded in the distribution of portamento types. The classification results suggest that mid-level and high-level features can facilitate the recognition of technique type in local-level. Major components to improve the classification are identified. Challenging cases are further discussed.

5.1 Data extraction

In playing technique detection, three types of data are derived from audio recordings. For the first and second types of data, we acquire the spectrogram using CQT, and obtain the pitch salience function using Crepe following the procedure stated in Section 4.1. For the third type of data, we further apply dynamic programming to eliminate the

---
[3] https://www.music-ir.org/mirex/wiki/2019:Audio_Key_Detection

spikes in the estimated pitch contour. Given the output of Crepe $X \in \mathbb{R}^{K \times N}$, the pitch salience function at time s_i is $X[:, s_i]$, K is the number of frequency bins, and N is the number of frames, the pitch contour $\mathcal{S} := \{s_i\}_{i=1}^{N}$ is extracted with the following objective function:

$$\mathcal{S}^* = \arg\max_{\mathcal{S}} \sum_{i=1}^{N} X[:, s_i] - \lambda \sum_{i=2}^{N} |s_{i+1} - s_i|. \quad (1)$$

Equation (1) can be solved with dynamic programming [27]. The seconds term of (1) is to enhance the smoothness of the extracted pitch contours. The parameter λ controls the smoothness of the pitch contour, and in this work we set $\lambda = 10^{-3}$. This facilitates the processing of those guqin historical recordings with relatively low audio quality.

Six types of playing techniques are labelled for the pitch variation movement in the left hand: *none* (such as 直按 and 散音), *vibrato* (such as 吟 and 猱), *upward portamento* (such as 綽 and 上), *downward protamento* (such as 注 and 下), *inverted-U protamento* (such as 進復 and 撞), and *U protamento* (such as 退復 and 豆). The three types of data mentioned above are then divided into event segments according to the onset and duration annotations, and each segment corresponds to one of the six technique types. Figure 3 illustrates the featured pitch contour and examples of event segments in 6 technique types. The segments are then treated with padding, resulting the the input segment size of cqt (156, 60), pitch salience function (361, 60), and pitch contour (361,).

In order to investigate that how the meta-, structural music features connect with technique implements in local-level, we further extract 62 high-level and mid-level features including 2 *music-level features* (mode, # of event in music), 12 *event-level features* (event duration, event index, note degree, importance ranking of degree, also 8 descriptive statistical indicators of pitch contour (the mean, maximum, minimum, range, standard deviation, skewness, kurtosis, the time difference between the maximum and minimum)). The first six features (the two features in music-level and the first four features in event-level) are obtained from the prediction of our mode detection model as described in section 4. The eight descriptive indicators are extracted from dynamic programming pitch contour. Since the critical traits of each technique type may appear in different parts of the overall pitch contour, we also extract *intra-event-level features* by computing the eight descriptive statistical indicators for six intra-event pitch con-
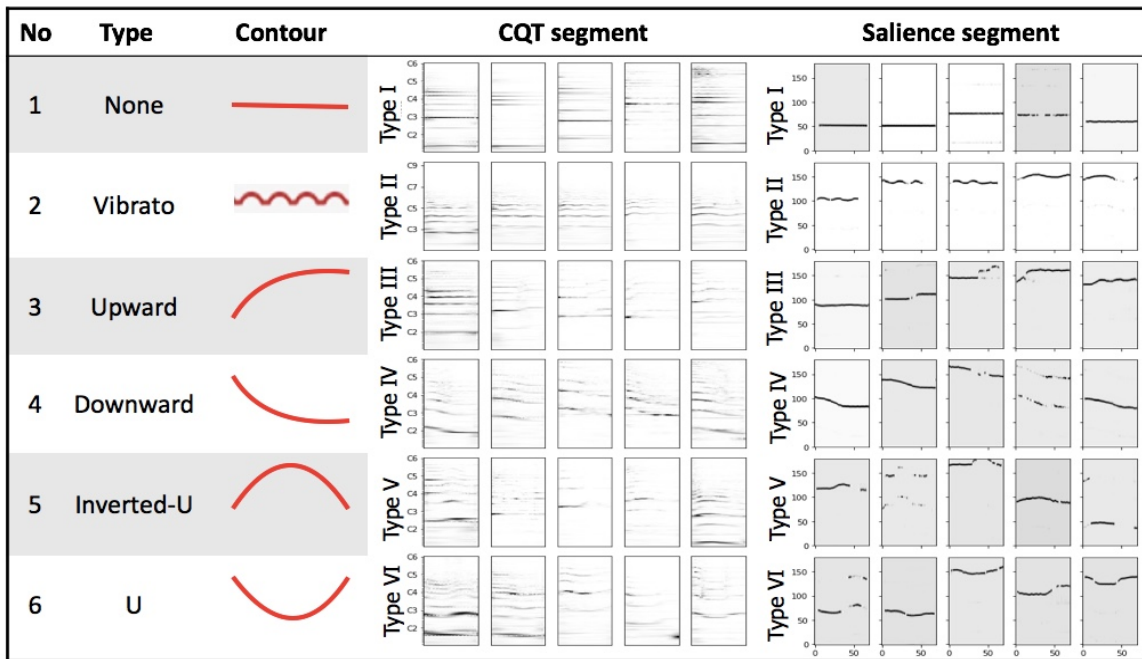
Figure 3. The CQT (middle) and pitch salience segments (right) corresponding to the six technique types of guqin.

Data type	CQT	Salience	Contour
Frame-level only	0.743	0.845	0.842
With mid-, high-level	0.840	0.839	0.842

Table 3. The average accuracy of the playing technique detection task using various local and high-level features.

tour per event. To define the range of intra-event pitch contour, each segment is resampled to the mean length of all segments (0.7 seconds), and the resampled contour is divided by six hop windows (window size = 0.2 seconds, hop = 0.1 seconds), resulting 48 intra-event-level features for per musical event (8 indicators x 6 windows).

5.2 Playing technique classification

The playing technique is classified by a 10-layer neural network, which is composed of 2 convolutional layers (kernel = 3 x 3, stride = 1, fmap = 32), 1 self-attention layer (kernel = 1 x 1), and then followed by 3 convolutional layers (kernel = 3 x 3, stride = 1, fmap = 32), 3 fully-connected layers (neuron = 512), and a softmax output layer. The attention layer is constructed based on [28] to further investigate which high- and mid- level elements affect the results of technique classification in local level. The framework is implemented using Tensorflow. The training process is carried out by minimizing the cross-entropy between the model output and the one-hot label using Adam Optimizer with the learning rate of 10^{-5}. The dataset and the code to implement the model will be released afterward.

We design six experimental settings to examine the interaction between the global and local features. Three types of frame-level data: CQT, pitch salience function, and pitch contour with dynamic programming processing are inputted into the network respectively. They are then associated with high-level and mid-level features derived from the prediction from the previous mode detection and labelled annotations, resulting six experimental settings in total (as shown in Table 3). For each setting, 10-fold cross-validation is performed (roughly 1800 seconds of audio recordings for training, and 200 seconds for testing).

5.3 Results and discussion

The statistical analysis shows that the distribution of portamento types reflects the high-level mode structure. As shown in Figure 4, except the type 1 technique without any pitch variation, type 3 portamento possesses highest occurrence ratio compared to other portamento types for important degrees in mode 1 (degree 1, 5, 2), and similar tendency can also be observed in other modes. This provides empirical evidence for the theoretical basis of guqin performing convention, where musicians frequently apply type 3 (upward) portamento to emphasize highlighted notes, and on the contrary, they tend to implement type 4 (downward) portamento to decorate trivial notes [16].

For the the results of technique classification, as can be seen in Table 3, the mid-level and high-level features contribute to the improvement of CQT data, but not for other two data types. This outcome may owing to the fact that the statistical descriptors of pitch contour are already contained in mid-level and high-level features, and such information of pitch contour may provide extra guidance for technique classification.

In order to further investigate that which are the decisive high-level and mid-level features to affect the technique classification, we follow the procedure in [28] and plot the self-attention with the feature map outputted from the neural network for 62 high- and mid- level features. All the values are normalized to the range between 0 and 1 for comparison. Figure 5 presents the attention map for the
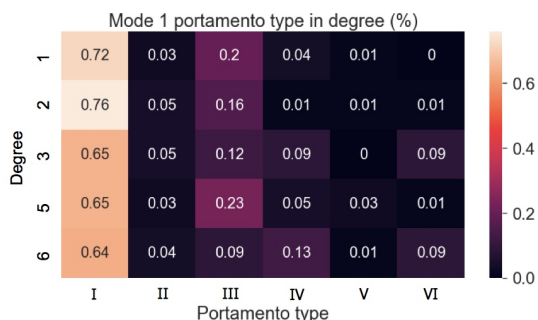
Figure 4. The statistics of the annotated portamento types (x-axis) in note degrees (y-axis) for mode 1.
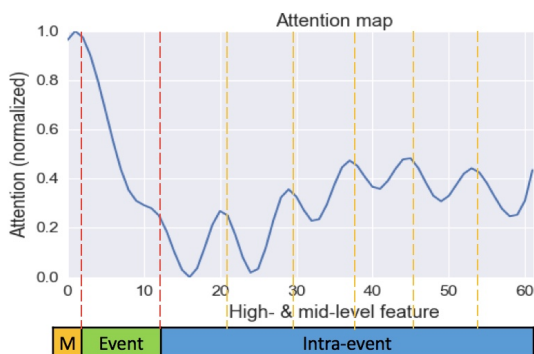

Figure 5. The attention map with feature map for 62 high- and mid- level features: music, event, and intra-event features. The red dotted lines indicate different feature types; the yellow dotted lines indicate features from 6 windows.
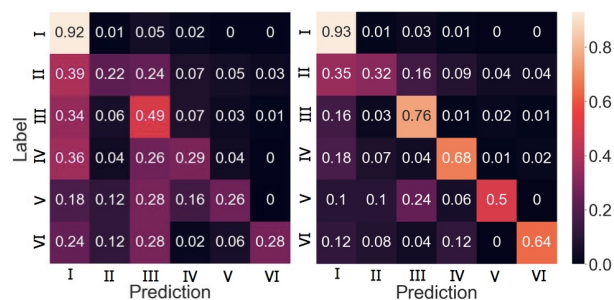

Figure 6. The confusion matrix of playing technique detection with CQT (left) and with CQT and high-, mid-level features (right).
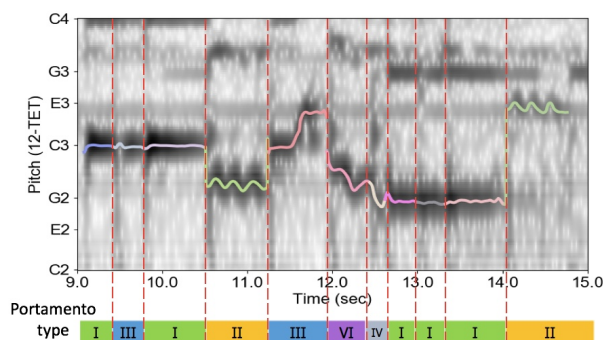

Figure 7. The example of CQT (the background figure) and pitch contour segments (color curves) for different portamento types (color blocks at the bottom).

CQT and mid-, high-level features combination. As can be observed in the graph, the music-level features receive the most attention in the classification model (mean value = 0.982), compared to the event-level features (mean value = 0.500) and intra-event-level features (mean value = 0.293). Moreover, in intra-event features, an obvious peak occurs in each window, which corresponds to a specific feature: the distance between the maximum and minimum value within the window. Above observations suggest that high-level features (music and event features) are more effective to facilitate technique classification. And in mid-level features (intra-event features), the distance between the maximum and minimum value is the most adequate element to represent the pitch contour constitution.

To analyze individual portamento types in depth, it is noted in the confusion matrix (Figure 6) that types II to IV (especially type II, vibrato) are easily confused with type 1 (no variation) portamento. It is an unexpected result contradicting the studies on other instruments [10, 29], where vibrato can be easily identified. The inter-type confusion is mainly caused by the data imbalance in guqin performance, in which type I occurs much more frequently than other types (see Figure 4). In addition, we further examine the spectrogram and pitch contours for different portamento types (Figure 7), and notice that the pitch contours of type 1 portamento are not straight lines as expected, but exhibit unstable and irregular pitch drift, which can be easily confused with weak type 2 portamento, partic-

ularly when only the spectral features are analyzed. Furthermore, the two type 3 portamento segments in this example display diverse manners. The divergent behaviour within one single portamento type can rise the difficulty for accurate classification. The insufficient quality of historical recordings may also blur the distinction between portamento types and add additional challenges for the task.

6. CONCLUDING REMARKS

In this paper, we design and compile a new dataset consulting the theoretical basis of guqin performance. Mode detection is performed on the collected dataset, and playing technique classification are conducted using neural network. The results indicate that the hierarchical construction is crucial for mode detection, and the high-level and mid-level features contribute to improving the playing technique classification task.

This work verifies the conventional theory by empirical observations, in which statistical analysis confirmed the solid connection between the mode structure and the arrangement of playing technique. This study highlights the joint-relationship between the global tonal structure and the local distribution of playing technique, as well as bridges the knowledge gap between the music theory and performance practice. The findings in this study contribute insights for constructing an auto-evaluation system for music performance, or an educational system for musicians and music listeners.

7. REFERENCES

[1] T. Wang, *Qin Zhi (The Purpose of Qin* 琴旨, *1744/1746). Si Ku Quan Shu (Wenyuange Edition, 1784) Vol. 220.* Taipei: The Commercial Press, 1986.

[2] L. W. Kong and T. Lee, "Automatic key partition based on tonal organization information of classical music," in *Proceedings of the the 15th International Society for Music Information Retrieval Conference*, 2014.

[3] F. Korzeniowski and G. Widmer, "Genre-agnostic key classification with convolutional neural networks," in *Proceedings of the the 19th International Society for Music Information Retrieval Conference*, 2018.

[4] C. Weiß and M. Schaab, "On the impact of key detection performance for identifying classical music styles," in *Proceedings of the the 16th International Society for Music Information Retrieval Conference*, 2015.

[5] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Raga recognition based on pitch distribution methods," *Journal of New Music Research*, vol. 41, no. 4, pp. 337–350, 2012.

[6] K. K. Ganguli, S. Gulati, X. Serra, and P. Rao, "Data-driven exploration of melodic structures in hindustani music," in *Proceedings of the the 17th International Society for Music Information Retrieval Conference*, 2016.

[7] T. Nuttall, M. García-Casado, V. Núñez-Tarifa, R. C. Repetto, and X. Serra, "Contributing to new musicological theories with computational methods: the case of centonization in arab-andalusian music," in *Proceedings of the the 20th International Society for Music Information Retrieval Conference*, 2019.

[8] N. Pretto, B. Bozkurt, R. C. Repetto, and X. Serra, "Nawba recognition for arab-andalusian music using templates from music scores," in *Proceedings of the the 15th International Sound and Music Computing Conference*, 2018.

[9] F. Scherbaum, M. Muller, and S. Rosenzweig, "Analysis of the tbilisi state conservatory recordings of artem erkomaishvili in 1966," in *Proceedings of the the 7th International Workshop on Folk Music Analysis*, 2017.

[10] Y.-P. Chen, L. Su, and Y.-H. Yang, "Electric guitar playing technique detection in real-world recordings based on f0 sequence pattern recognition," in *Proceedings of the 16th International Society for Music Information Retrieval Conference*, 2015.

[11] L. Su, L.-F. Yu, and Y.-H. Yang, "Sparse cepstral and phase codes for guitar playing technique classification," in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, 2014.

[12] J. Abeber, H. Lukashevich, and G. Schuller, "Feature-based extraction of plucking and expression styles of the electric bass guitar," in *Proceedings of IEEE Acoustics, Speech and Signal Processing (ICASSP)*, 2010.

[13] P.-C. Li, L. Su, Y.-H. Yang, and A. W. Y. Su, "Analysis of expressive musical terms in violin using score-informed and expression-based audio features," in *Proceedings of the 16th International Society for Music Information Retrieval Conference*, 2015.

[14] C.-C. Shih, P.-C. Li, Y.-J. Lin, Y.-L. Wang, A. W. Y. Su, L. Su, and Y.-H. Yang, "Analysis and synthesis of the violin playing style of heifetz and oistrakh," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, 2017.

[15] J.-I. Liang, *The Event of 'Cadenza' in Liu Shui of Pan-Chuan Qin School and Ping Sha of Mei An Qin School: A Thesis on the Great Changes of Qin Music at the Transition from the Old to the New.* Taipei: Master's thesis, Taipei National University of the Arts, 2012.

[16] ——, *The Characteristics of Modal Structure in Qin Compositions of the Florid-Inflections Style Since the Late Ming Dynasty: Analysis and Application of the 'Ti-Yong' Theory in Wang Tan's Qin Zhi.* Taipei: PhD dissertation, Taipei National University of the Arts, 2018.

[17] S.-Y. Xu, *Da Huan Ge Qin Pu (Da Huan Ge Pavilion Anthology of Qin Music* 大還閣琴譜, *1673). Reprinted in Qin Gu Ji Cheng (Anthology of Qin Music* 琴曲集成*) Vol. 10.* Bejing: Chung Hwa Book Company, 2010.

[18] S.-J. Chen, *Qin Xue Chu Jin (Introduction to the Scholarship of Qin* 琴學初津, *1902). Reprinted in Qin Gu Ji Cheng (Anthology of Qin Music* 琴曲集成*) Vol. 28.* Bejing: Chung Hwa Book Company, 2010.

[19] M.-G. Gu, *Qin Xue Bei Yao (Required Essentials of Qin Scholarship* 琴學備要*).* Shanghai: Shanghai Music Publishing House, 2009.

[20] W.-C. Chou, "Single tones as musical entities: An approach to structured deviations in tonal characteristics," *American Society of University Composers*, vol. 3, pp. 86–97, 1970.

[21] C. H. Lee, *Asian music.* Yang-Chih Book, 2015.

[22] W. G. Wu, "The modal system of qin and its verification 1," *Musicology in China*, vol. 1, pp. 5–30, 1997.

[23] ——, "The modal system of qin and its verification 2," *Musicology in China*, vol. 2, pp. 88–109, 1997.

[24] P. Guo, *Inimitable Sound and Treasures: The Collection of Historical Audio and Video Tracks from Guqin Legends by GuoPeng (*絕響: 國鵬輯近世琴人音像遺珍*).*

[25] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, vol. 8, 2015.

[26] J. W. Kim, J. Salamon, P. Li, and P. Bello, "Crepe: A convolutional representation for pitch estimation," in *Proceedings of IEEE Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[27] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2009.

[28] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proceedings of the 36th International Conference on Machine Learning*, 2019.

[29] C. Wang, E. Benetos, V. Lostanlen, and E. Chew, "Adaptive time-frequency scattering for periodic modulation recognition in music signals," in *Proceedings of the 20th International Society for Music Information Retrieval Conference*, 2019.