

Design Ideas for an AI-Driven Pipa Playing Pitch and Rhythm Judgment System

Junxi Li^{1*}

¹ Guangzhou Xinhua University, Guangzhou, 523132, China

* lijunxi@xhsysu.edu.cn

<https://doi.org/10.70695/IAAI202602A9>

Abstract

Addresses the problem in traditional pipa teaching where pitch and rhythm feedback relies heavily on teachers' on-site listening and is difficult to continuously record during practice. Based on the principle of "assisting teaching rather than replacing evaluation", this paper proposes an AI-driven system for determining pipa pitch and rhythm. The system uses pipa solo audio, standard scores, teacher review comments, and practice feedback requests as primary data, forming a closed loop of "data acquisition—preprocessing—intelligent judgment—score alignment—evidence feedback—continuous correction". The design focuses on summarizing insights from recent core journal research on traditional music datasets, automatic music annotation, time-frequency attention, performance evaluation, and human-computer collaboration. It also considers the characteristics of pipa techniques such as tremolo, strumming, pushing and pulling, vibrato, harmonics, and ornamentation, proposing the system structure, key modules, data processing, deployment route, and feedback boundaries. The system output is not centered on experimental scores but rather on measure location, error evidence, teacher review, and practice suggestions, providing a design reference for the construction of intelligent teaching systems for traditional Chinese plucked string instruments. In terms of pitch accuracy, the accuracy of instrument manufacturing and playing actions are further incorporated into the system logic: the former corresponds to the process calibration of frets, phase and key positions, while the latter corresponds to the left hand's pressing position, string contact angle and string pressing stability, and can be used with the help of machine vision recognition to form action evidence.

Keywords Artificial Intelligence; Pipa Performance; Pitch Accuracy; Rhythm Determination; Music education ; Machine Vision

1 Introduction

Pipa playing incorporates various techniques such as tremolo, strumming, pushing and pulling, vibrato, harmonics, and ornamentation. Its pitch variations, reverberation patterns, and timing organization differ from the relatively standardized musical signals of piano and vocal instruments. In classroom teaching, teachers can rely on experience to judge pitch and rhythm issues, but in large-class instruction, after-class practice, and review sessions, subtle deviations are often difficult to consistently record. The development of intelligent audio perception and music information retrieval provides new tools for assisting in the identification of pipa playing processes; however, system construction still needs to be redesigned based on the acoustic characteristics of traditional plucked string instruments, the context of musical scores, and teaching feedback methods.

Pitch accuracy determination requires distinguishing between the instrument's objective accuracy and the accuracy of the playing action. The former is mainly determined by the manufacturing process, similar to piano tuning and key positioning. For the pipa, only when the frets, phase, string spacing, and fingering contact points are accurate can the standard pitch relationships have a stable foundation. The latter is mainly determined by the player's left-hand fingering technique. When there are deviations in the fingering position, fingertip force, string contact angle, and shifting path, even if the instrument itself is accurately manufactured, it will result in pitch deviations, underpitch deviations, or pitch drift. Therefore, intelligent systems should not rely solely on fundamental frequency estimation but should also incorporate left-hand fingering position recognition, hand shape stability, and musical note pitch mapping into auxiliary evaluations.

Research in core journals over the past three years shows that music artificial intelligence has expanded from single audio recognition to a comprehensive discussion of traditional music datasets, automatic music annotation, cross-modal transcription, and human-computer collaboration. Research on traditional music annotation datasets emphasizes the joint organization of audio, score, technique labels, and performance context [3]; research on pipa performance event recognition suggests that the acoustic form and technique of ethnic plucked instruments need to be included in the recognition process [5]; reviews on automatic music annotation and time-frequency attention studies indicate that pitch, onset, and duration should not be treated separately, but should be judged in combination with frequency structure, time structure, and score context [2,6-8].

Philosophical reflections on music AI, human-machine relationships, and emotional transmission studies further suggest that intelligent systems need to properly handle the boundary between technical efficiency and artistic subjectivity [1,4,12]. Performance evaluation reviews emphasize that a clear boundary needs to be maintained between technical accuracy, musical expression, and evaluation consistency. Intelligent systems are suitable for providing technical evidence and recording practice processes, but should not replace teachers' comprehensive judgment on style, emotion, and artistic treatment [9]. Audio Transformer reviews provide methodological insights for modeling short-term onset, beat boundaries, and long-term dependencies [10]. Model bias studies also remind us that system design should not rely solely on general models or single scenarios, but should retain mechanisms for teacher review, scenario adaptation, and continuous correction [11].

Based on the above research, the key to intelligent judgment of traditional plucked string instruments lies not in pursuing a single model score, but in establishing a teaching support process that is collectable, interpretable, and verifiable. This design positions the system as a "performance evidence processing tool", transforming students' intonation tendencies, onset positions, note values, vibrato, and redundant notes from their practice into discussionable teaching materials through audio acquisition, score analysis, event candidate generation, error prompts, and teacher review.

2 Design Overview

2.1 Application Scenarios and Development Goals

The system is mainly designed for three scenarios: ethnic instrumental music classes, social training, and after-school self-study. The classroom end uses the teacher's tablet or classroom workstation as the main entry point, the student end uses mobile recording devices and web feedback interfaces as the main entry points, and the teacher review end is used to view common problems in the class and individual practice tracks. The system does not replace the teacher's comprehensive evaluation of musical expression, style and emotion, but only provides traceable prompts around pitch tendency, rhythm position and score consistency, so that teachers can more quickly locate measures, notes and recurring technical problems [9]. The design goals are shown in Table 1.

Table 1. Key design considerations for an intelligent judgment system for pipa performance

| project | Design content | illustrate |
|-------------------------|---|---|
| Application Entry | Classroom terminal, student terminal, and teacher review terminal | Balancing classroom instruction, after-class exercises, and review management |
| Source of information | Pipa solo, fixed practice tasks, and free performance excerpts | Maintain task comparability and authentic practice conditions |
| Data Composition | Performance audio, sheet music index, anonymous number, teacher comments | Form verifiable teaching materials |
| Judgment content | Pitch tendency, onset position, duration relationship, omission of notes, redundant notes | Feedback on sections and technique segments |
| Construction principles | Support for instruction, evidence review, teacher verification, and privacy protection | Avoid replacing art evaluation with system scores |

2.2 Data Acquisition and Engineering Conditions

Data collection employs a combination of "fixed practice tasks and free performance excerpts". Fixed tasks facilitate comparison of differences among different students using the same piece and technique, while free excerpts preserve tempo variations, dynamic changes, and embellishments from actual practice. Data is primarily stored as audio clips of performances, sheet music indexes, anonymous student IDs, and teacher review comments; no identity images are collected, and no information that can be used to track individual identities is output. Before entering the system, silent clipping, loudness adjustment, environmental noise checks, and segmentation are performed to ensure that subsequent identification is based on relatively clear teaching materials (Figure 1).

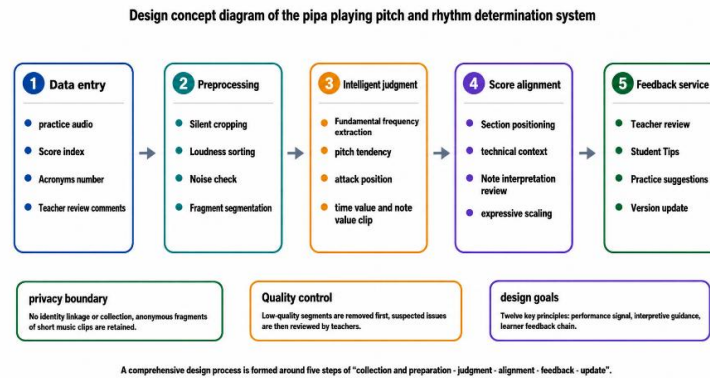


Fig. 1. Design concept diagram of an AI-driven pipa playing pitch and rhythm determination system

The system consists of five parts: data entry, preprocessing, intelligent judgment, score alignment, and feedback service. The data entry part associates performance audio, score index, and practice tasks; the preprocessing part performs silent trimming, loudness adjustment, noise checking, and segmentation; the intelligent judgment part generates pitch, attack, and note value candidates based on time-frequency cues; the score alignment part places the candidate events back into the specific measure, note, and technique context; and the feedback service transforms problems into practice prompts that teachers can review and students can understand. The focus of the output is not on generating MIDI files or single scores, but on forming traceable teaching evidence [3,5,8].

2.3 Difficulties in Judging Errors in Pipa Performance

The pipa's sound is characterized by strong transients, short sustains, diverse finger noise, and dense attacking sounds. Tremolo easily creates continuous, subtle attacking sounds, strumming produces dense spectral peak changes, and push-pull and vibrato cause continuous pitch fluctuations. If only the fundamental frequency is relied upon for estimation, decorative pitch fluctuations may be misinterpreted as intonation problems; if only abrupt spectral peak changes are relied upon, aftereffects and finger noise may be misinterpreted as new attacking sounds. Therefore, the system needs to simultaneously consider local transients, long beats, score positions, and technical context, and reserve "uncertain segments" for teacher review.

3 System Design and Key Methods

3.1 Audio Preprocessing and Time-Frequency Characterization

Preprocessing first completes loudness unification, silence trimming, and environmental noise checks to ensure basic comparability of audio samples collected from different devices and classrooms. Then, time-frequency cues such as CQT and Mel spectra are extracted: CQT is better suited to representing semitone structures and harmonic relationships, while Mel spectra are better suited to preserving short-term energy changes. Parallel input of these two types of cues helps the system simultaneously focus on pitch tendency and rhythmic transients. Score parsing, based on MusicXML, MIDI, or manually

transcribed scores, transforms standard scores into reference sequences containing measures, notes, beats, durations, string positions, and fingering cues, providing a basis for subsequent alignment [2,8].

3.2 Hierarchical Time-Frequency Determination Framework

As shown in Table 2, the system adopts a hierarchical time-frequency attention framework as its core design. The frequency direction is used to focus on the differences between string positions, overtones, resonances, and ornaments; the time direction is used to focus on the onset, aftertones, tremolo continuity, and rhythmic organization; and the score constraints are used to limit candidate events to reasonable measures and note ranges. This framework draws on the research ideas of time-frequency attention, cross-modal music transcription, and audio Transformer, but in the context of the pipa, it places greater emphasis on technical context, teacher review, and teaching feedback [6-7,10].

Table 2. Key modules and functions of the hierarchical time-frequency determination framework

| Module | Enter clues | Output clues | Design Role |
|--------------------------------|---|--|---|
| Frequency sensing layer | CQT/Mel and other spectral clues | Differences in string position, harmonics, and ornaments | Support for pitch tendency judgment |
| Time-aware layer | Frame-level audio clips and beat positions | Attack, Aftershock, and Timing Relationship | Support rhythm organization judgment |
| Musical score constraint layer | Measures, time signatures, fingering, and technique markings | Candidate event range | Reduce misjudgment of free velocity and ornamentation |
| Multiple head prompt layer | The fused audio and sheet music clues | Pitch, attack, duration, and review suggestions | Generate hierarchical feedback information |
| Teacher review layer | Suspected problems and low-confidence segments | Corrections and version history | Establish a sustainable renewal mechanism |
| Visual sound recognition layer | Left fingertip position, fingering area, and string contact angle | Deviance of finger placement and hand shape stability | Supplementing evidence of pitch and movement |

In terms of attention design, the system does not process the entire audio segment evenly. Instead, it focuses attention on the vicinity of candidate notes based on the position in the score and the practice task. The frequency perception layer first distinguishes acoustic cues such as string position, overtones, and finger noise; the time perception layer then analyzes the onset, aftertone, and beat relationships; and the score constraint layer is responsible for determining whether candidate events conform to the standard musical phrase structure. This structure helps reduce misjudgments caused by free tempo, ornaments, and local noise.

The event output employs a multi-headed prompting approach, avoiding a single "right" or "wrong" conclusion. Instead, it provides separate suggestions for pitch tendency, onset position, duration stability, and review. For segments with clear clues, the system can directly generate practice prompts; segments with dense strumming, obvious vibrato, or significant background interference are marked as requiring teacher review. This approach preserves the efficiency of AI processing while avoiding mechanical judgments that overlook reasonable differences in musical expression.

3.3 When Making Specific Judgments on Pitch, Rhythm, and Feedback Rules

The system first confirms the instrument's manufacturing and calibration basis, including whether the fret positions, phase relationships, string spacing, and standard pitch reference are stable; when the instrument itself has accurate fingering positions, it then proceeds to the deviation analysis at the playing action level.

At the manufacturing level, pitch accuracy can be understood as the stability of the objective pitch positions of an instrument. After the pipa is manufactured and calibrated, each pitch position should form a relatively fixed pitch reference, and its judgment logic is similar to that of the stability of the piano key pitch. The system can establish a mapping relationship between the standard score pitch, the instrument tuning, and the corresponding pitch positions, first eliminating the influence of tuning

deviation, fret errors, and local pitch instability on the recognition results, and avoiding simply attributing instrument manufacturing problems to student playing errors.

At the level of playing technique, the left-hand fingering position is a direct factor in pitch accuracy. Fingertips landing too far forward or backward, insufficient string pressure, incorrect string contact angle, and unstable return to position after shifting all contribute to a deviation from the target pitch. The system can incorporate machine vision recognition beyond audio judgment. Through key hand point detection, fingertip trajectory tracking, and pitch region matching, it can determine whether the left hand is in the correct position and correlate "audio deviation" with "movement deviation", providing teachers with clearer error correction criteria. For pitch-related prompts, a correlation between "instrument calibration verification—left-hand fingering position—audio deviation result" can be added, enabling teachers to determine whether the problem stems from the instrument's objective pitch, the student's fingering technique, or a combination of both.

Pitch feedback uses a grading system of "too high", "too low", "basically stable", and "needs review", retaining the corresponding measure and note position. For push-pull, vibrato, and ornamentation, the system doesn't consider all pitch fluctuations as errors; instead, it assesses their appropriateness by considering notation, technique hints, and the relationship to preceding and following phrases. Teachers can view more complete candidate evidence, while students primarily receive brief hints and practice directions.

Rhythmic feedback is organized around the timing of the onset, duration extension and contraction, continuity, missed notes, and redundant notes. Slow practice focuses more on the clarity of the onset and the integrity of the duration of each note, while fast tremolos and strumming emphasize continuity and rhythmic stability. Free-speed passages require distinguishing between expressive extension and contraction and genuine rhythmic loss of control. Systematic feedback should avoid single-threshold error judgments and instead provide tiered prompts based on the structure of the score and the teacher's review comments.

The comprehensive tips are presented in a textual format, organized in the order of "Problem Location - Possible Cause - Practice Suggestions - Review Entry". The problem location points to a measure, note, or technique segment; possible causes include unstable hand position, incorrect finger placement, uneven finger density during tremolo, and insufficient strumming continuity; practice suggestions correspond to slow practice, string-by-string practice, metronome assistance, and teacher demonstration and listening practice.

To ensure usability for instruction, the feedback interface needs to differentiate between three categories of information: "clear problems", "potential problems", and "performance-based processing". Clear problems are suitable for direct inclusion in the student's practice list, potential problems are added to the teacher's review list, and performance-based processing is retained as material for classroom discussion. Through this classification, the system shifts from a scoring tool to a tool for organizing teaching evidence, which better reflects the actual process of learning pipa performance.

4 Construction Route and Implementation Ideas

4.1 Data Organization and Route Marking

The data organization is based on teaching tasks, rather than simply expanding the sample size. Each segment needs to be associated with the score number, practice task, tempo requirement, technique type, and teacher review comments. The organization process can follow the route of "task definition - audio organization - teacher review - data archiving - module adaptation - deployment and update": first, clarify the piece segments and technique objectives, then complete the audio cleaning and quality check, then the teacher reviews the pitch tendency, rhythmic position, and technical context, and finally the stable materials are archived and used for system iteration (Figure 2).

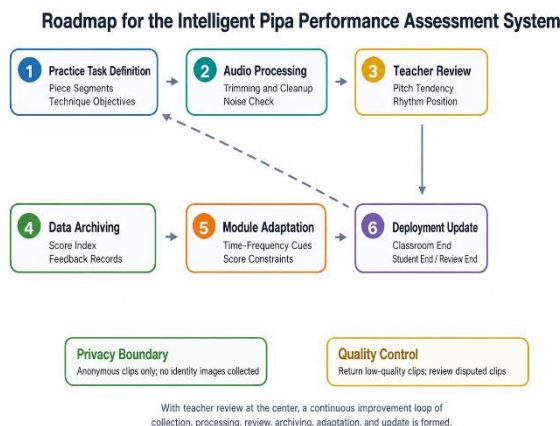


Fig. 2. Roadmap for pipa performance data compilation, teacher review, and system update

Table 3. References, lessons learned, and design transformations from core journal research in the past three years

| Reference Field | Main content | Design inspiration | Corresponding steps |
|--|---|--|----------------------------------|
| Research on the Boundaries of Music Artificial Intelligence | Human-machine relationship, artistic subjectivity and emotional transmission [1,4,12] | The system is clearly defined to only handle the compilation of technical evidence and provide supplementary feedback. | System positioning |
| Research on Traditional Music Data and Annotations | Audio, sheet music, and technique tags work together [3,5] | Establish anonymous fragments, sheet music indexes, and teacher review materials. | Data collection and organization |
| Automatic Music Annotation and Time-Frequency Attention Research | Attack, pitch, duration, and time-frequency structure analysis [2,6-8,10] | Place pitch and rhythm within a unified judgment framework | Intelligent judgment |
| Performance Evaluation Research | Technical accuracy, musical expression, and scoring boundaries [9] | Transform feedback into tiered practice suggestions | Teaching feedback |
| Model Bias and Scene Adaptation Research | Acoustic conditions, style, and data bias [11] | Retain teacher review, scenario adaptation, and continuous updates. | Deployment and maintenance |

4.2 Module Configuration and Deployment Route

The system development does not end with a one-time model configuration, but rather proceeds along the route of "basic recognition — scene adaptation — teacher review — version update". The basic recognition module is responsible for handling common etudes and standard tempo segments; the scene adaptation module focuses on noise differences in recordings from ordinary classrooms, practice rooms, and mobile devices; the teacher review module is responsible for correcting suspected problems and supplementing technique explanations; and the version update module transforms stable review opinions into the basis for subsequent system optimization. For pitch-related feedback, the teacher can simultaneously view the matching of the left-hand key point trajectory with the target pitch area, while the student's end receives action-oriented prompts such as "press the note slightly forward", "return to position too low after shifting", and "fingertip landing point needs to be close to the target pitch".

The deployment employs a combination of edge-based initial assessment and cloud-based review. Recording, preprocessing, and initial prompts are completed on mobile devices or classroom workstations, while the cloud primarily handles batch review by teachers, data management, and cross-class question aggregation. The system does not save complete, long-term recordings; it only retains

segmented and anonymized teaching segments, sheet music indexes, and feedback records. This deployment approach balances classroom responsiveness, privacy protection, and subsequent maintenance.

4.3 Determine the Feedback Output Format

The output is divided into teacher and student versions. The teacher version presents common problems encountered in the class, difficulties in the piece, review records, and practice suggestions, facilitating differentiated instruction. The student version only presents necessary measure prompts, intonation tendencies, rhythmic issues, and practice directions, avoiding turning system feedback into complex technical reports. Feedback language should closely resemble classroom language, such as "the start of this measure is slightly rushed", "the fingering in this section is uneven", and "the return to position after the push-pull technique needs further stabilization", enabling students to directly translate this into practice actions.

5 Design References and Application Paths

5.1 References and lessons learned from relevant research

Research in core journals over the past three years has provided five types of insights for intelligent judgment of pipa performance: First, research on the boundaries of music artificial intelligence suggests that the system needs to adhere to auxiliary positioning [1,4,12]; second, research on the construction of traditional music data suggests that the system should simultaneously organize audio, scores and technique tags [3,5]; third, research on automatic music annotation and time-frequency attention suggests that pitch and rhythm judgment need to pay attention to frequency structure and time structure at the same time [2,6-8,10]; fourth, research on performance evaluation suggests that feedback language should serve the teaching process rather than replace artistic judgment [9]; and fifth, research on model bias suggests that the system needs to retain scene adaptation and teacher review mechanisms [11].

These connections illustrate that general music transcription models cannot be directly applied to pipa lessons. The system design needs to connect "audio events that technology can recognize" with "teaching evidence that teachers need to view", placing audio events, score locations, technical context, and practice suggestions within the same teaching chain. Only in this way can intelligent judgment information be transformed from technical output into actionable teaching feedback.

5.2 Teaching Application Approach

Classroom applications can proceed along the path of "pre-class preparation — in-class data collection— instant prompts—teacher review—post-class practice". Before class, the teacher selects the repertoire, excerpts, and technical objectives; during class, students record practice excerpts and generate initial prompts; the teacher demonstrates common problems based on system prompts; after class, students practice slowly and listen to sections repeatedly based on the section prompts. Throughout this process, the system handles recording, location, and prompting functions, while the teacher remains responsible for musical expression, stylistic treatment, and learning strategy arrangement.

From a learning perspective, beginners need clear onset and basic pitch cues, intermediate students need cues on fingerpicking, strumming, and duration consistency, while advanced students need cues on tempo, phrasing, and technique stability. The system can break down feedback for the same piece into basic, technical, and expressive layers, allowing students of different levels to find practice paths suitable for their individual stages. Further improvements to the pitch module should focus on the synergy between machine vision and audio recognition: audio is responsible for determining actual pitch deviations, while vision is responsible for identifying left-hand finger placement and technique stability; both types of evidence work together to help teachers determine the causes of pitch problems.

5.3 System Boundaries and Improvement Directions

The system's boundaries need to be clearly defined: audio models can help identify traceable technical problems, but they cannot fully determine musical aesthetics, emotional expression, or genre

style. For free tempo, dynamics, timbre selection, and personal expression, the system can only provide auxiliary clues; the final judgment still needs to return to teacher demonstrations, student listening and identification, and classroom discussions.

Further improvements can be made in three directions. First, expand the notation and technique tags to provide clearer explanations for special techniques such as push-pull, vibrato, harmonics, and strumming. Second, optimize the teacher review interface to create a continuous link between system prompts, teacher explanations, and student practice records. Third, explore the combination of audio and motion information while adhering to privacy regulations to provide more teaching clues for issues such as hand shape, string contact position, and force application. Related bias studies and emotion transmission studies also suggest that continuous updates should simultaneously consider scene differences, subjective experience, and feedback ethics [11-12].

6 Conclusion

This design constructs an AI-driven system for determining pitch and rhythm in pipa playing, forming a closed loop from audio acquisition, score analysis, time-frequency representation, candidate event determination, score alignment to teaching feedback. The system is positioned as a teaching aid, with its core task being to organize verifiable evidence rather than replace teacher evaluation. Among these, pitch assessment should be supported by "instrument manufacturing accuracy-left-hand fingering technique-fusion of audio and visual evidence", so that the system can not only indicate pitch deviations but also explain the action reasons for the deviations.

Based on the characteristics of the pipa, such as strong transients, short sustained notes, dense tremolo and rich ornamentation, the system adopts a hierarchical time-frequency judgment framework, which combines frequency clues, time clues, musical constraints and technical context to avoid simply judging ornamentation as an error.

The reference value of core journal research in the past three years is mainly reflected in the boundaries of music artificial intelligence, the construction of traditional music data, automatic music annotation, time-frequency attention, performance evaluation, and model bias control. The system design needs to transform these methods into a collection process, review process, and feedback language that can be used in the classroom, rather than just focusing on the comparison of model indicators.

Future development should continue to improve the technique labeling, teacher review, and tiered feedback mechanisms, and explore the integration of audio and motion information while adhering to privacy compliance. Through continuous refinement, the system can better support pipa classroom teaching, after-class practice, and periodic learning reviews.

Acknowledgement

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Huang, Z. (2023). Philosophical reflection on music artificial intelligence. *Journal of Central Conservatory of Music*, (3), 9-21.
2. Zhang, R., Wang, H., Liu, L., et al. (2023). A review of research on automatic music annotation and classification methods. *Computer Science and Exploration*, 17(6), 1225-1248.
3. Zhang, Y., & Sun, M. (2024). Research on the construction of traditional music annotation dataset for artificial intelligence-oriented—and on the construction and application of multimodal erhu annotation dataset. *Journal of Central Conservatory of Music*, (2), 66-83.
4. Ma, S. (2024). Artificial intelligence in music composition: Reflections on intervention methods and human-machine relationship. *Journal of Central Conservatory of Music*, (2), 84-100.

5. Wang, Y., Li, X., Zhang, Y., et al. (2025). Towards automatic expressive pipa music transcription using morphological analysis of photoelectric signals. *Sensors*, 25(5), 1361.
6. Wang, Q., Liu, M., Bao, C., et al. (2024). Harmonic-aware frequency and time attention for automatic piano transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 3492-3506.
7. Gu, X., Ou, L., Zeng, W., et al. (2024). Automatic lyric transcription and automatic music transcription from multimodal singing. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20(7), 1-29.
8. Müller, M., & Chiu, C. Y. (2024). A basic tutorial on novelty and activation functions for music signal processing. *Transactions of the International Society for Music Information Retrieval*, 7(1), 179-194.
9. Moura, N., Dias, P., Verissimo, L., et al. (2024). Solo music performance assessment criteria: A systematic review. *Frontiers in Psychology*, 15, 1467434.
10. Zaman, K., Li, K., Sah, M., et al. (2025). Transformers and audio detection tasks: An overview. *Digital Signal Processing*, 158, 104956.
11. Marták, L. S., Hu, P., & Widmer, G. (2026). Sound and music biases in deep music transcription models: A systematic analysis. *EURASIP Journal on Audio, Speech, and Music Processing*, 2026, 5.
12. Xie, D. (2025). Embedding, alienation and decryption: A critical examination of the emotional logic of artificial intelligence music dissemination. *Journal of Southwest University for Nationalities (Humanities and Social Sciences Edition)*, 46(11), 156-164.

Biographies

1. **Junxi Li** Master, currently a full-time teacher at the Department of Music, Guangzhou Xinhua University, with a research focus on Pipa performance art.

基於人工智能驅動的琵琶演奏音準與節奏判定系統設計思路

李君熙¹

¹廣州新華學院，廣州，中國，523132

摘要：針對傳統琵琶教學中音準與節奏反饋主要依賴教師現場聽辨、課後練習過程難以連續記錄的問題，圍繞「輔助教學而非替代評價」的原則，提出一種人工智能驅動的琵琶演奏音準與節奏判定系統設計思路。系統以琵琶獨奏音頻、標準曲譜、教師複覈意見和練習反饋需求為主要資料，形成「採集—預處理—智能判定—曲譜對齊—證據反饋—持續修正」的閉環。設計上側重綜述近三年核心期刊研究中關於傳統音樂數據集、音樂自動標註、時頻注意力、演奏評價和人機協同的啓示，並結合琵琶輪指、掃弦、推拉、吟揉、泛音和裝飾音等技法特點，提出系統結構、關鍵模塊、資料整理、部署路線和反饋邊界。系統輸出不以實驗分數為中心，而以小節定位、錯誤證據、教師複覈和練習建議為核心，為民族彈撥樂器智能教學系統建設提供設計參考。在音準判定方面，進一步把樂器製作準度與演奏動作準度納入系統邏輯：前者對應品位、相位和按鍵位置的工藝校準，後者對應左手按音位置、觸弦角度和按弦穩定性，並可藉助機器視覺識別形成動作證據。

關鍵詞：人工智能；琵琶演奏；音準判定；節奏判定；音樂教育；機器視覺

1. 李君熙，碩士，現任廣州新華學院音樂學院專任琵琶教師，廣東文藝職業藝術學院外聘琵琶/阮教師，廣東民族管絃樂學會理事，廣東民族管樂學會彈撥樂協會會員，廣東省民族管絃樂學會嶺南音樂普及專業委員會理事，廣州市越秀區音樂家協會嶺南音樂推廣專業委員會副主任，廣州市越秀區音樂家協會會員，廣州市民族樂團琵琶首席，廣東廣播電視臺珠江頻道藝人。曾多次參與國內外大型比賽與演出活動，榮獲多項省級、國家級獎項。