Development and Usability Evaluation of an Intelligent Personalized Erhu Pitch and Rhythm Learning System

Gwo-Haur HWANG[a*], Ping-Tsung TSAI[b], Jenn-Kaie LAIN[b] & Shiuan-Han HUANG[c]
[a]*Graduate School of Intelligent Data Science, National Yunlin University of Science and Technology, Taiwan*
[b]*Department of Electronic Engineering, National Yunlin University of Science and Technology, Taiwan*
[c]*Chinese Music Department, Tainan National University of the Arts, Taiwan*
*ghhwang0424@gmail.com

Abstract: Pitch accuracy and rhythm errors are the most common problems encountered by erhu beginners. The traditional way of learning the erhu involves using a tuner and a metronome to improve these two issues. However, for erhu beginners, without the assistance of a teacher, it is often difficult to address these problems effectively. Therefore, this study has developed an "Intelligent Personalized Erhu Pitch and Rhythm Learning System" to help erhu learners learn to pitch and rhythm anytime and anywhere without the presence of a teacher while providing professional feedback. This system included pitch and rhythm training features. In addition to automatically generating sheet music in the MML (Music Macro Language) format, it allows the erhu learners to play along with the sheet music. After each performance, the system provides learners with expert-like comments on pitch and rhythm, respectively. The system also keeps a record of learners' practice sessions and provides appropriate encouragement and guidance based on their learning progress by creating an intelligent and personalized learning scaffold. After the development of the system, this study conducted a usability evaluation and investigated the relationship between usability and erhu playing experience. The results showed that the average for all usability dimensions was above 4.6, indicating that the system is acceptable for erhu learners. Moreover, learners with more erhu playing experience showed higher acceptance of the system compared to those with less experience in erhu.

Keywords: Intelligent, personalized, erhu, pitch learning, rhythm learning, usability evaluation

1. Introduction

1.1 Background

Erhu is popular among the Chinese community because of its cultural identity and expressiveness. Erhu lessons can be conducted in both one-on-one and group classes. Although the quality and progress of group classes may not be as good as one-on-one teaching lessons, most amateur learners prefer group classes because of the low price, the ease of making friends, and the relaxed learning environment.

1.2 Motivation and Purpose

Erhu is not easy to learn because it needs proficient skills to control the tone color, pitch, and rhythm. Beginners will need to rely on a teacher for a long time to master those skills. However, it takes more time for a teacher to demonstrate and explain in group classes, which makes it impossible to correct every problem with rhythm or pitch (Hwang, Chen, Sung, & Kuo, 2015).

In traditional erhu group teaching, students practice pitch and rhythm with the teacher and using a tuner and metronome to correct pitch and rhythm. Learners are often unable to remember the mistakes they have made. The metronome stabilizes the rhythm, but it is less likely to be used as a tool for learning rhythm. Therefore, this study has developed an "Intelligent Personalized Erhu Pitch and Rhythm Learning System" to address the above problems of erhu learning to improve erhu learners' pitch and rhythm.

2. Literature Review

2.1 Technology-Assisted Pitch Learning

The sound of a bowed string instrument is controlled by pressing the strings, which is difficult for beginners because they need to control fingering, bowing posture, and develop pitch listening skills. Therefore, Wang, Wang, Chen, Chang, and Chen (2012, July) developed a violin pitch detection system that provides scales and arpeggio scores as teaching materials. Violin learners can choose different rhythms to practice according to their level. Unlike the violin, the erhu does not have a fingerboard to press, and all pitches need to be pressed with the left hand floating in the air.

Previous research has developed two erhu learning systems to improve finger position, pitch (Kikukawa, Ishihara, & Soga, 2013), and bowing (Kikukawa & Soga, 2014) using magnetic position sensors and display interface. Although these systems for erhu beginners are accompanied by a visualized interface, the device is still too complicated for those who need to practice constantly. Each time they wear the device, they may become less willing to learn. Therefore, there is a need for a more convenient erhu learning system that allows learners to learn anytime and anywhere without wearing any equipment, as long as they have a tablet.

2.2 Technology-Assisted Rhythm Learning

Another important element of playing an instrument is rhythm. It is very important to have a correct and stable rhythm when playing music. However, beginners are often influenced by their own senses, resulting in a fast or slow rhythm that affects their overall performance. Therefore, learners usually use a metronome to stabilize their instruments.

Many studies have used body postures and beats to improve learners' rhythms, either indirectly or directly. For example, Fonteles and Rodrigues (2021) proposed a Kinect-based conducting system that allows students to imitate a conductor and thus train rhythm stability. At the same time, the system proved to be effective, simple and flexible. As another example, Dos Santos, Loke, Yacef and Martinez-Maldonado (2022) proposed a method for modeling rhythmic movements through smartphones. The smartphone's motion sensor enhanced the dance teachers' assessment of rhythm-related dance skills. These two studies were conducted to generate and evaluate music-related motion features and to enhance rhythm evaluation through dance teacher examination.

Although these systems are helpful for learners' rhythm, they are not suitable for erhu learners because the erhu requires the coordination of left and right hands. Therefore, we developed the “Intelligent Personalized Erhu Pitch and Rhythm Learning System” to help learners learn how to play the erhu.

2.3 Intelligent Personalization Systems

In order to achieve personalized teaching and learning, an assisted learning system should not only provide learning resources and a learning environment for learners, but also actively interact with them to understand their learning situation and imitate the teaching style of real teachers to provide different teaching materials for different users. For example, Desmarais and Baker (2012) suggested that intelligent learning systems should embed learner models that effectively deal with uncertainty and partial evidence. This helps students know more about their learning progress. In addition, Zhe (2021) designed an Intelligent Guided Learning System (IGLS) model for music sight-reading, which not only combines the characteristics of the music sight-singing subject in teaching contents and teaching methods but also proposes a difficulty feature-based score recommendation algorithm and the sight-singing scoring algorithm.

Therefore, the system developed in this study incorporated the concept of intelligent personalized learning, allowing learners to learn on their own at any time and from anywhere, as well as incorporating comments from real erhu teachers and recording their learning history to provide a personalized learning guide.

3. System Design

3.1 System Architecture

The system architecture diagram is shown in the left of Figure 1. After entering the system, the learners can first tune and then choose to practice the pitch or rhythm. The system will generate the corresponding sheet music for the learners. It will analyze and provide a score and suggestions for each pitch after the learner completes their performance according to the instructions. The results will be recorded for learners to check at any time.

3.2 Tuning Mode

The system has a built-in tuner for learners to tune before practicing. The screenshot of the built-in tuner is shown in the right of Figure 1. Before using the tuner, learners can choose to adjust the A4 frequency. Although the international standard for A4 is 440 Hz, some performers or orchestras will adjust the A4 frequency from 436 Hz to 444 Hz in order to achieve the best performance. When the learner adjusts the A4 frequency of the tuner of this system, the system will update the A4 frequency of the practice mode at the same time.

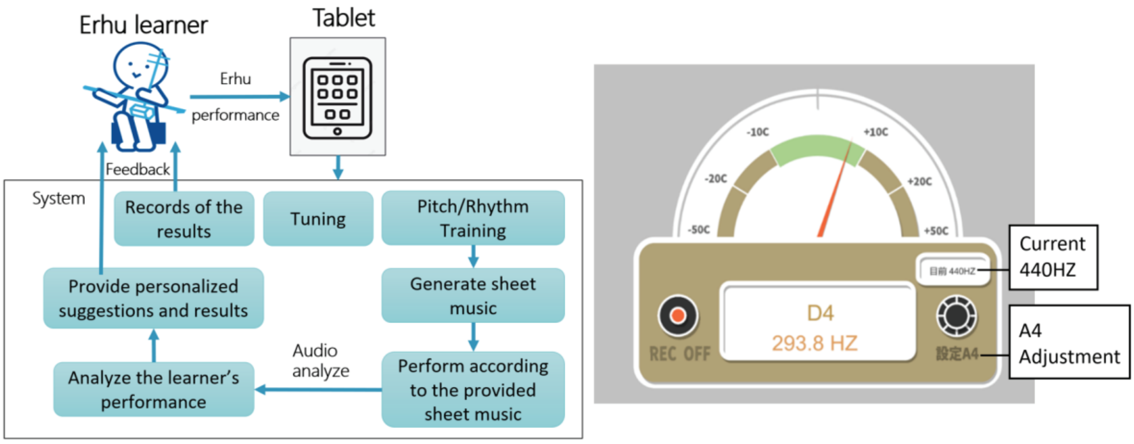

Figure 1. System Architecture Diagram and Screenshot of the Built-in Tuner.

3.3 Pitch Training Diagnostic Mode

Since a semitone is equal to 100 cents, a reasonable range for a pitch is plus or minus 50 cents, i.e., a frequency exceeding plus or minus 50 cents is considered out of tune. The human ear can distinguish at least 25 cents (Peretz & Hyde, 2003), so when comparing, the exact pitch can be set at about 20–25 cents, and when the detected frequency falls within plus or minus 20 cents after the cents calculation, it can be judged as the correct pitch. The detailed phonetic score judging criteria are shown in Table 1.

Table 1. *Tone Division Judgment Standard*

Judgment	Judgment Criteria
Correct	$\lvert cent \rvert < 20$
Too high	$20 < \lvert cent \rvert < 50$ & $cent > 0$
Too low	$20 < \lvert cent \rvert < 50$ & $cent < 0$
Out of tune	$\lvert cent \rvert \geqslant 50$
Not played	$f = 0$

The pitch training mode, as shown on the left of Figure 2, allows learners to choose the level they want to practice at. There are descriptions of the topic below each level's button so that learners can choose the appropriate level according to their ability.

The sheet music produced by the pitch training mode includes scale exercises based on the commonly used keys of the erhu (D, G, C, F, A, ^{b}B) and three positions (two octaves). The note will be marked during the performance to guide the learner. The correct frequency waveform (black) and detected frequency waveform (red) will be displayed on the top of the interface so that the learners can instantly check their performance status, as shown in the right of Figure 2.

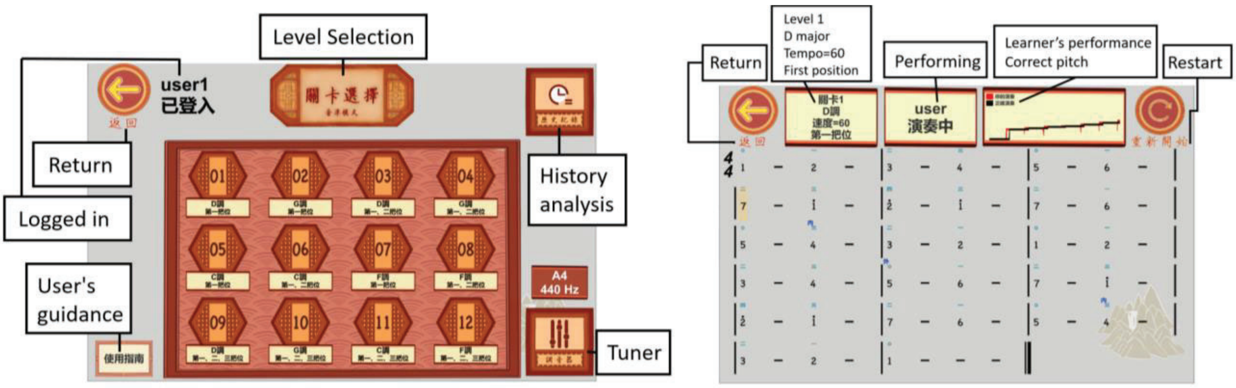

Figure 2. Screenshot of Level Selection and During Performance.

After playing, the system will analyze the learner's performance and display the detection results for each note below the corresponding note, as shown in Figure 3. There are four judgment levels, each represented by a color, and each judgment level is scored according to the judgment level. In the performance results, learners can click on a note on the score to enter the individual diagnosis interface. The individual diagnosis interface of the pitch mode (Figure 4) shows information about the performance judgment (correct, high, low, severely high, and severely low), the judgment problem, and suggestions for correction.

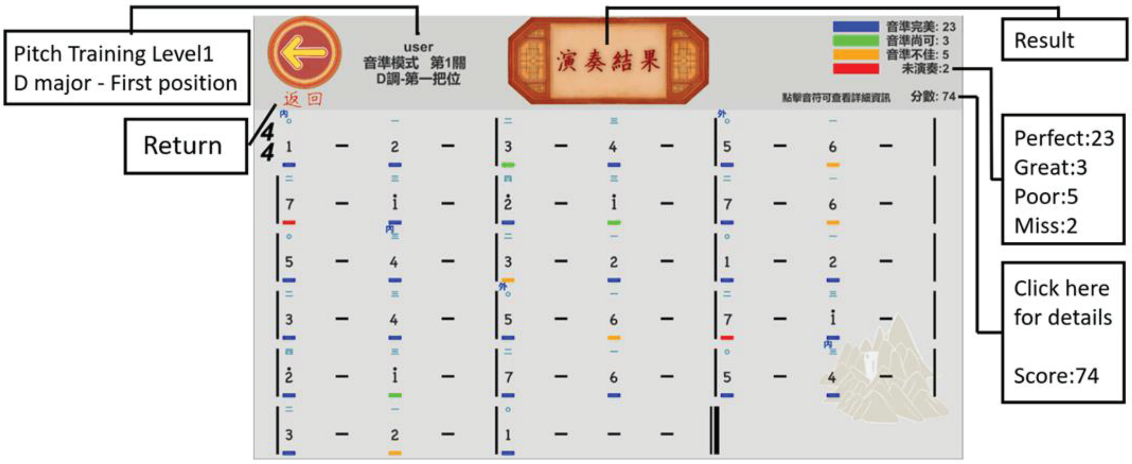

Figure 3. Screenshot of the End of Performance.

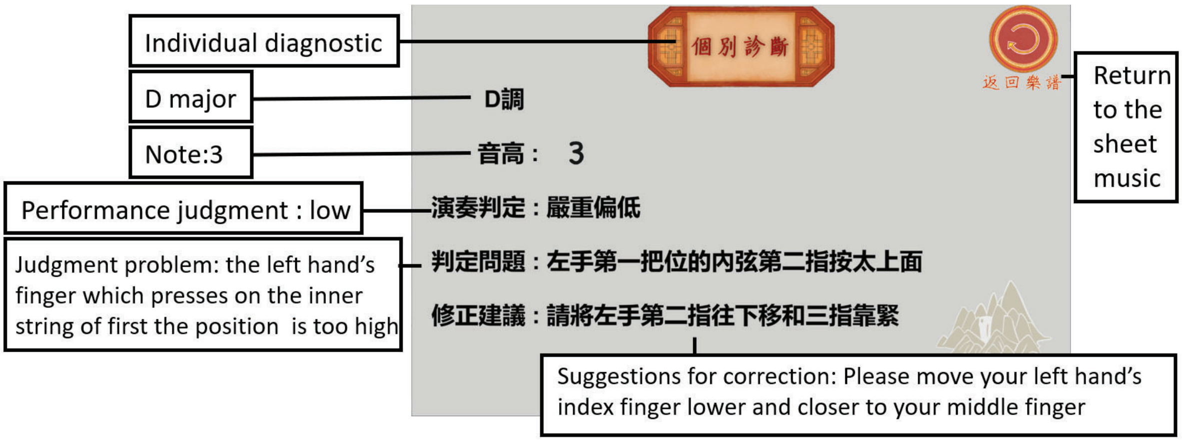

Figure 4. Individual Diagnostic Screenshot of the Pitch Training Diagnostic Mode.

3.4 Rhythm Training Diagnostic Mode

The criteria for detecting rhythm is to compare the detected start time, end time and duration with the correct start time, end time and duration, respectively. The threshold value for both start and end times is TH, and the threshold value for the duration is 20% of the correct time. There are four judgments for start time, end time, and duration. The four judgments for start time are "not played", "beat rush", "beat drop", and "correct"; the four judgments for end time are "not played", "beat rush", "beat drag", and "correct"; and the four judgments for duration are "not played", "too short", "too long", and "correct". The criteria for calculating the rhythm score are shown in Table 2.

Table 2. *Rhythm Score Calculation Criteria*

Status	Judgment criteria	Score	Weight
Pitch	Exact string	1	1
	Wrong string	0	
Start time	Time Difference < TH	1	1
	Time Difference > TH	0	
End time	Time Difference< TH	1	1
	Time Difference > TH	0	
Duration	Time Difference < 20%	1	1
	Time Difference > 20%	0	
Rhythm coverage rate	Coverage rate >80%	2	4
	50%< Coverage rate <80%	1	
	Coverage rate <50%	0	
The number of points obtained = Score x Weight Rhythm score = (Total number of points obtained / 12) x 100			

The rhythm training diagnostic mode generates sheet music in different rhythms, including the common erhu rhythms (4 beats, 3 beats, 2 beats, 1 beat, half beat, etc.), and the rhythm practice is in D major. The notes are composed of the inner string and the outer string played together. The individual diagnostic interface of the Rhythm mode (Figure 5) shows information about the correctness of the empty strings (inner and outer strings), the error time of the entry point, the error time of the breakpoint, the error of the playing length, the rhythm coverage, and the rhythm score, with the criteria shown in Table 3.

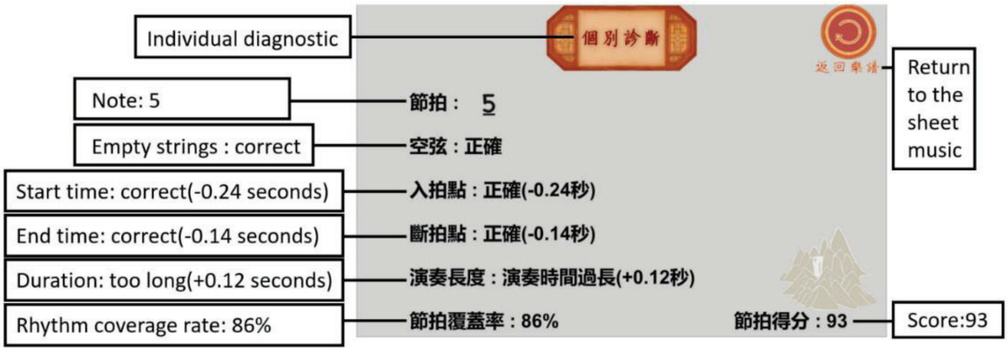

Figure 5. Rhythm Mode Individual Diagnosis Interface.

Table 3. *Judgment Levels and Scoring Criteria*

Color	Sound level mode	Rhythm mode	Score	Weight
Blue	|Cent| ≤ 20	Rhythm Score ≥ 90	1	3
Green	20 < |Cent| ≤ 50	60 ≤ Rhythm Score < 90	0.5	2
Orange	|Cent|>50	0 < Rhythm Score < 60	0	1
red	Not played	Rhythm Score = 0	0	0

3.5 Intelligent Personalized History Analysis and Guidance Model

After each performance, the system will record the results of the performance. In the history analysis interface, each record will show the serial number, user name, selected practice mode, level, score, and performance date. The history analysis interface is shown in Figure 6.

After playing, the learner can click on the button in the upper right corner for history analysis to view past performances and compare the accuracy of each month's performance.

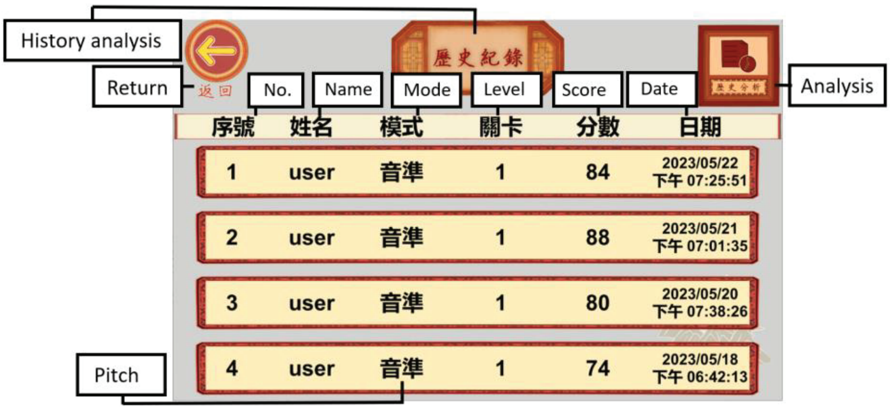

Figure 6. History Analysis.

The system will analyze the learner's learning history by month, mainly divided into three analysis modes: "Analysis of this month", "Comparison of specified months" and "Trend of performance accuracy". Learners can click on the upper-right corner button to check their past record in the History Analysis and use the drop-down menus to select the items they want to view:

Analysis of this month: The analysis for this month is shown on the left of Figure 7. The system first collects all the records of the learner's performance at the specified level in this month and then calculates the accuracy (percentage) of each item in the level (pitch training mode is the pitch that appears in the track, rhythm training mode is the beat that appears in the track) according to the weights in Table 3. The system calculates the sum of the weights of each item in the month and the total number of items and then divides the sum and the total number by 3 (the weight is 3) to calculate the accuracy of the item in the month. At last, the accuracy of every item will be presented in a bar chart.

Specified month comparison: The schematic diagram of the specified month comparison is shown on the right of Figure 7. The calculation method is the same as the "analysis of this month". Learners can compare whether they have progressed or regressed in playing the item this month with previous months.

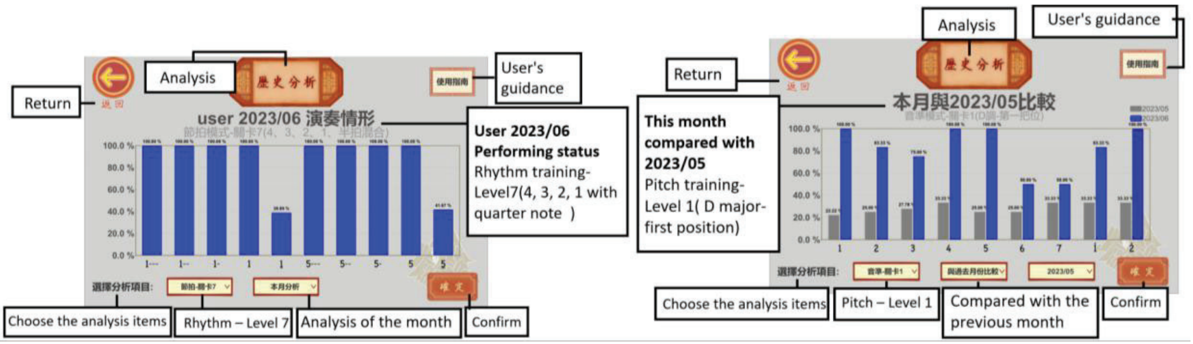

Figure 7. Analysis of the Month (Rhythm Mode) and Comparison of Specified Months.

Performance Accuracy Trend: The performance accuracy trend graph is shown in Figure 8. This mode analyzes the average accuracy of the learner for each month at the specified level. The system lists the months in which the learner played the specified level from the oldest to the newest, calculates the average accuracy for each month, and then presents the average accuracy trend for the specified level as a line graph.

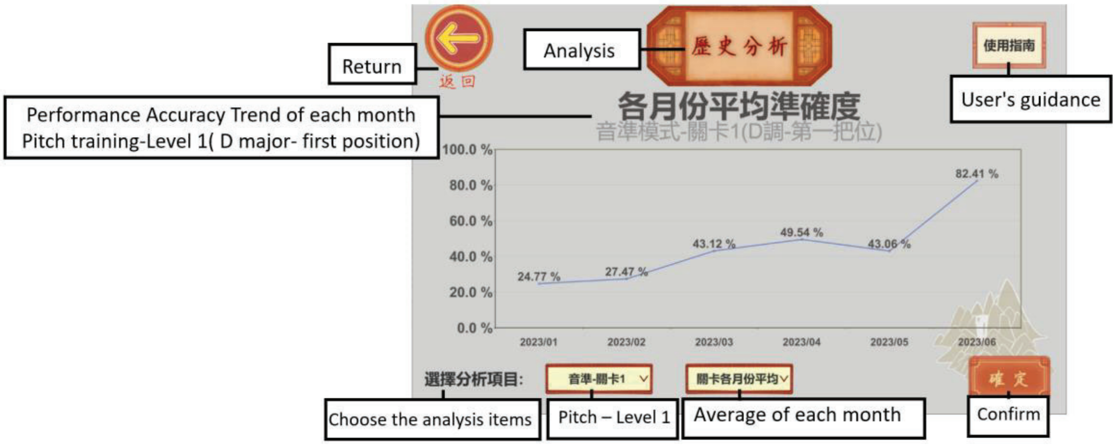

Figure 8. Performance Accuracy Trend.

4. Experimental Results and Analysis

4.1 System Availability Evaluation

After the system was developed, in order to investigate whether the interface and operation of the system are suitable for erhu learners, erhu students and teachers were invited to try out and operate the "Intelligent Personalized Erhu Pitch and Rhythm Learning System" developed by this study. They need to fill out the usability evaluation questionnaire, and an interview will be conducted to find out their thoughts after operating the system.

The usability evaluation questionnaire used in this study was planned and designed according to the 10 user interface usability evaluation criteria of the system proposed by Nielsen (1994) and Nielsen (1994, April) and was designed using Likert (1932) five-point scales. Each question has 5 options: "Strongly agree (5 points)", "Agree (4 points)", " Neither agree nor disagree (3 points)", "Disagree (2 points)" and "Strongly disagree (1 point) ". Each dimension has three questions according to the system design, and the entire questionnaire has a total of 30 questions.

4.2 Experimental Subjects

The subjects of this study were 34 erhu learners from a music classroom in central Taiwan and erhu majors from the Department of Chinese Music at a university of the art in southern Taiwan. The age distribution ranges from 9 to 74 years old, from 1 to 10 hours of erhu practice per week, and from less than 1 year to more than 10 years of erhu playing experience.

4.3 Experimental Process

The experimental procedure for system usability evaluation in this study is shown in Figure 9. Before the experiment starts, the experimental subjects will be informed of the experimental precautions and asked to fill out the informed consent form for 10 minutes. After that, the experimental subjects were introduced to the functions of the system and explained the system's operation procedure for about 10 minutes. Each experimental subject performed the system operation according to the learning sheet. The experimental subjects will be allowed to use the tuner of the system to tune the erhu for about 5 minutes after the explanation of the system operation. Then, the experimental subjects will be allowed to practice pitch and rhythm, respectively. They can check the sheet music for each mode and choose two of them to practice for about 15 minutes. After all the tracks are practiced, the experimental subjects will be allowed to view the performance records and analysis of the performance for about 5 minutes. Lastly, a 15-minute usability evaluation questionnaire and a post-use interview were conducted to find out what they thought of the system. The experiments are shown in Figure 10.

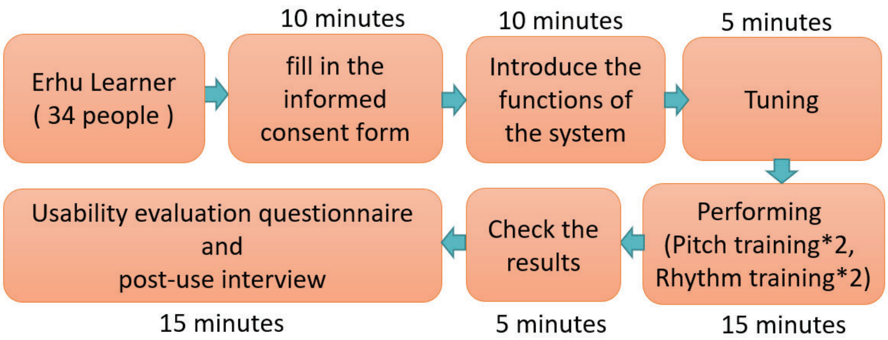

Figure 9. The Experimental Procedure.

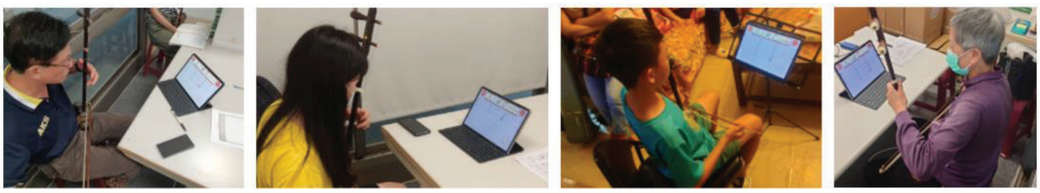

Figure 10. Usability Evaluation Experiments.

4.4 Data Collection and Analysis

After all the subjects had completed the experiment and filled out the questionnaires, all the data were analyzed using SPSS statistical software. First, a reliability analysis was carried out. The results showed that the Cronbach's Alpha values of each facet were greater than 0.7 (between 0.755 and 0.906), indicating that all questionnaires reached an acceptable level.

After analyzing the reliability of the questionnaire, we then analyze the average of each dimension. The usability evaluation results can be seen in Table 4. The average of all dimensions is greater than 4.6. The results show that the system developed by this research is suitable for erhu learners. Among them, the top three with higher averages are H4, H7 and H5.

Table 4. *Result of Usability Evaluation*

Usability evaluation dimensions	Count	Min	Max	Ave	SD	Ranking
H1：Visibility of system status	34	3.0	5.0	4.690	0.478	5
H2：Match between system and the real world	34	3.0	5.0	4.618	0.520	10
H3：User control and freedom	34	3.0	5.0	4.710	0.448	4
H4：Consistency and standards	34	4.0	5.0	**4.770**	0.374	**1**
H5：Error prevention	34	3.0	5.0	**4.740**	0.433	**3**
H6：Recognition rather than recall	34	3.0	5.0	4.670	0.519	8
H7：Flexibility and efficiency of use	34	3.0	5.0	**4.750**	0.421	**2**
H8：Aesthetic and minimalist design	34	3.0	5.0	4.680	0.522	6
H9：Help users recognize, diagnose, and recover from errors	34	4.0	5.0	4.670	0.512	9
H10：Help and documentation	34	3.0	5.0	4.680	0.522	7
Average of the dimensions	34	3.4	5.0	4.696	0.437	

In Table 5, erhu learners with more erhu playing experience had significantly higher averages in H6, H7, H8, and H9 than erhu learners with less experience, which means they have higher acceptance of the system.

Table 5. *Impact of Erhu Playing Experience on Usability Evaluation*

Usability evaluation dimensions	Erhu playing experience	Count	Ave	SD	*t*
H1：Visibility of system status	>=8 years	16	4.792	0.382	1.221
	<=7 years	18	4.593	0.543	
H2：Match between system and the real world	>=8 years	16	4.688	0.412	0.734
	<=7 years	18	4.556	0.605	
H3：User control and freedom	>=8 years	16	4.854	0.344	1.891
	<=7 years	18	4.574	0.496	
H4：Consistency and standards	>=8 years	16	4.854	0.344	1.179
	<=7 years	18	4.704	0.394	
H5：Error prevention	>=8 years	16	4.875	0.295	1.894
	<=7 years	18	4.611	0.502	
H6：Recognition rather than recall	>=8 years	16	4.875	0.342	**2.425***
	<=7 years	18	4.481	0.585	
H7：Flexibility and efficiency of use	>=8 years	16	4.917	0.258	**2.315***
	<=7 years	18	4.611	0.488	
H8：Aesthetic and minimalist design	>=8 years	16	4.896	0.264	**2.590***
	<=7 years	18	4.481	0.618	
H9：Help users recognize, diagnose, and recover from errors	>=8 years	16	4.875	0.363	**2.450***
	<=7 years	18	4.481	0.563	
H10：Help and documentation	>=8 years	16	4.854	0.344	2.010
	<=7 years	18	4.519	0.607	
Average of the dimensions	>=8 years	16	4.848	0.307	**2.048***
	<=7 years	18	4.561	0.497	

*p<.05

5. Conclusion and Future Prospects

5.1 Conclusion

In this study, an "Intelligent Personalized Erhu Pitch and Rhythm Learning System" was developed using Unity. We evaluated the usability of the system and investigated the impact of erhu playing experience on the usability evaluation. According to the experimental result, the average of each usability dimension is higher than 4.6, which means the system is acceptable for erhu learners. Learners who have more erhu playing experience have higher acceptance of the system. In other words, the more experienced erhu learners will be more receptive to this system than the less experienced erhu learners.

5.2 Future Works

After the usability evaluation experiment, post-experimental interviews were conducted with all subjects in order to find out what they really thought about the system developed in this study. In terms of advantages, we hope to add some functions such as tempo customization, voice control, etc. In terms of system planning, we hope to diversify the topics or allow users to decide their own sheet music so that the practice will not be limited by the system. In the future, we will continue to modify the system according to the suggestions above and improve the stability of the system. We expect to conduct more complete teaching experiments on the learning effectiveness of erhu learners to investigate whether the system can improve their accuracy and stability of pitch and rhythm. In addition, in the future, the influence of learners of different ages on the acceptance of this system and the learning effect can also be explored.

Acknowledgements

This research was funded by the National Science and Technology Council, Taiwan, R.O.C. with grant number NSTC 111-2410-H-224-004-MY2.

References

Desmarais, M. C., & Baker, R. S. D. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, *22*, 9-38.

Dos Santos, A. D. P., Loke, L., Yacef, K., & Martinez-Maldonado, R. (2022). Enriching teachers' assessments of rhythmic Forró dance skills by modelling motion sensor data. *International Journal of Human-Computer Studies*, *161*, 102776.

Fonteles, J. H., & Rodrigues, M. A. F. (2021). User experience in a kinect-based conducting system for visualization of musical structure. *Entertainment Computing*, *37*, 100399.

Hwang, G. H., Chen, B., Sung, C. W., & Kuo, T. H. (2015, November). To explore the effect of age on cognitive load when using social websites to conduct flipped classrooms on musical instrument performance teaching-Taking amateur erhu learners as an example. In *Proceedings of the 23rd International Conference on Computers in Education (ICCE 2015)*, Hangzhou, China (pp. 1-8).

Kikukawa, F., Ishihara, S., Soga, M., & Taki, H. (2013, May). Development of a learning environment for playing erhu by diagnosis and advice regarding finger position on strings. In *NIME* (Vol. 2013, pp. 271-276).

Kikukawa, F., Soga, M., & Taki, H. (2014). Development of a gesture learning environment for novices' erhu bow strokes. *Procedia Computer Science*, *35*, 1323-1332.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 1-55.

Nielsen, J. (1994). *Heuristic Evaluation. Usability Inspection Methods*. New York: John.

Nielsen, J. (1994, April). Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 152-158).

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362-367.

Wang, J. H., Wang, S. A., Chen, W. C., Chang, K. N., & Chen, H. Y. (2012, July). Real-time pitch training system for violin learners. In *2012 IEEE International Conference on Multimedia and Expo Workshops* (pp. 163-168). IEEE.

Zhe, T. (2021). Research on the model of music sight-singing guidance system based on artificial intelligence. *Complexity*, *2021*, 1-11. https://doi.org/10.1155/2021/3332216.