AEM JOURNAL.ORG

Research on the Design and Application of Immersive Teaching Environment for Guzheng Performance Based on Virtual Reality (VR)

Miao Wang
School of music, Nanjing University of the Arts, Nanjing 210013, Jiangsu, China
Corresponding author: Miao Wang (e-mail: 13247762931@163.com)

ABSTRACT Conventional guzheng teaching has long faced structural problems such as limited teaching space, delayed feedback in teacher-student interaction, and insufficient multisensory immersion. To address these issues, this study develops and applies an immersive teaching environment for guzheng performance based on virtual reality (VR), and systematically explains the technical architecture, interaction mechanism, and teaching process of its main modules. A complete teaching environment is constructed by integrating hand posture capture, audio-visual synchronization, virtual instrument modeling, real-time audiovisual feedback, and task-sequence-driven practice. From an engineering perspective, the system also relies on stable sensor signal acquisition, low-latency wireless interaction, and synchronized information transmission, which are important for maintaining immersion and interaction continuity in VR-based music education. An eight-week teaching intervention involving 30 beginner guzheng learners was conducted under a controlled experimental design. The results show that the experimental group significantly outperformed the control group in fingering standardization, with an average score 22.4 points higher; rhythmic stability, with the average deviation reduced by 54.4 ms; and intrinsic learning motivation, which increased by 37.5%. The proposed VR teaching environment can partially mitigate the increased cognitive load caused by multisensory digital instruction through adaptive feedback mechanisms, enhance motor retention, and provide empirical evidence and a technical route for the digital reconstruction of guzheng teaching and other instrumental music education scenarios supported by immersive interaction and wireless sensing technologies.

INDEX TERMS virtual reality, guzheng teaching, immersive learning, wireless sensing, audiovisual synchronization

I. INTRODUCTION

BEING one of the most representative traditional musical instruments in China, the playing techniques of the guzheng are complicated, requiring the learners to be highly precise in their fingers, coordinated, and expressive in music. The recent mainstream paradigm of oral transmission of information, however, by individuals or large groups, has always been subject to a clash of popularization and individualization. The students tend to form the wrong muscle memory as they are not able to compare their playing position with the minor variations of the reference model. It is also hard to correct the shape of the hand of every student, the angle of contact of the strings and the change of the force frame by frame in the limited time available in the classroom. This method which relies on a delayed auditory reception and visual observation is not efficient enough to support the increasing social learning requirements.

Thus, the main questions of this work are as follows: How to have a VR immersive environment that is in accordance with the actual needs of teaching guzheng performance? How will the most important technical modules (hand tracking, audiovisual feedback, and teaching sequence) within this environment be designed and combined? What is the effect of this VR teaching environment on the standardization of fingering, rhythmic accuracy and learning motivation in beginners compared to traditional teaching? The paper will offer a repeatable and testable way of digital transformation of teaching of the traditional Chinese musical instruments using technical solution design and empirical data.

II. RELATED WORK

Virtual reality (VR) technology provides a new technical path for building a visual and interactive immersive performance learning environment. Phakamach et al. [1] explored the potential application value of the metaverse in higher education and vocational education based on the analysis of educational technology trends. Roche et al. [2] pointed out that immersive learning effectively improves students' emotional investment and conceptual understanding depth under the modular teaching framework. Khamis et al. [3] found that immersive technology is most widely used in the fields of medicine, engineering, science education and teacher training, and generally shows a positive impact on spatial ability, procedural knowledge memory and problem-solving ability. Lowell and Yan [4] proposed that systems thinking can provide a holistic perspective for the design of IVR learning experience. Bizami et al. [5] summarized five innovative teaching principles for immersive blended learning.

Ng et al. [6] explored whether and how the use of music applications in online flipped classrooms can promote students' creative music production and their learning experience. Yin [7] analyzed teachers' cross-cultural strategies and professional growth when using MOOCs to teach guzheng. Huang et al. [8] analyzed the specific challenges encountered in the inheritance of guzheng in Shaanxi and proposed a comprehensive strategy that combines modern educational technology with social participation. Khlaif et al. [9] sorted out the opportunities and challenges of XR in engineering education. Lee and Wu [10] explored the impact of implementing microteaching in an IVR environment on teachers' emotions, teaching decisions and skills development. Furthermore, research in the field of VR-assisted fine motor skills has established that real-time visual scaffolding can accelerate early skill acquisition in violin and piano performance. Existing literature has demonstrated the potential of immersive technology in education from a macro perspective and has made preliminary explorations in the digitalization of music teaching. However, when focusing on the typical fine motor skill scenario of guzheng performance, it has failed to construct an immersive closed-loop training environment that includes adaptive multimodal feedback and cognitive load regulation. This study designs a VR immersive environment specifically customized for guzheng performance teaching and uses empirical data to test its actual effectiveness and underlying mechanism in improving the core performance skills of beginners.

III. METHODS

A. HAND POSTURE CAPTURE MODULE AND 3D MODELING SPECIFICATIONS

The basis of building a VR teaching system of Guzheng performance is the ability to capture the hands of the learner, particularly, the angle of the joints, the direction of the fingertips, and the weight exerted by the learner at the point of contact on the strings (thumb, index, middle, and ring fingers). In this experiment, a scheme is used which uses optical and inertial fusion tracking. The hardware layer uses the Leap Motion 2.0 controller that has a 150 degree field of view and can track finger displacement with sub-millimeter precision at distances of 25 to 60 cm of the sensor. These accuracy metrics were validated against a Vicon Vantage V5 optical motion capture system as the ground-truth reference. While consumer-grade evaluations often report 1-3 mm errors in dynamic environments, our system achieved a refined average fingertip error of 0.43 mm through a custom optical and inertial fusion algorithm and strict calibration of the sensor's mounting angle on the HMD. This controller is attached to the lower front edge of the head-mounted display (HTC VIVE Pro Eye), so that the field of view of the sensor will be directed to the gaze direction of the learner. The software layer obtains raw hand skeleton data via the Gemini driver of Ultraleap, producing 26 degrees of freedom hand posture parameters on a per-frame basis, such as the bending angles of the three phalanges of each finger, the lateral swing angle of the metacarpophalangeal joints, and the spatial rotation quaternion of the wrist. To map the captured free-playing movements to the string position model of the virtual guzheng, a strict spatial coordinate transformation standard needs to be established. To begin with, the guzheng has twenty-one strings that are characterized in the virtual scene. Each string has a special three-dimensional line segment of space, whose initial point is a fixed point on the front bridge and its final point is the respective point on the back bridge. The strings are lined in pitch order starting with lowest string (first string) to highest string (twenty-first string) at equal spacing in the x-axis with a spacing of 8 millimeters between the strings. Second, the conditions of identification of valid string-plucking events are specified: when the spatial coordinates of any fingertip (the fingertip of the index or middle finger in the model) satisfy three conditions, the system identifies it as a string-plucking event: the Euclidean distance between the point and the target string segment is less than 2.5 millimeters; The system samples positions and determines conditions every millisecond, and measures the time when each string-plucking event occurs, which finger is in contact with the string, where on the front bridge the touching point is, (measured in millimeters), and the instantaneous velocity of the touching point of the string.

To prevent VR system latency due to overly complex models when simplifying the hand, in the visualization and rendering of the 3D model of the hand, the hand model is represented by a layered simplified polygonal mesh, with the total number of faces being limited to less than 800 triangles, although preserving the visual characteristics between the finger joints. The model material and lighting options will be a semi-transparent outline of whiteness with an internal light blue halo to increase its visual identification with complex virtual backgrounds. Raw data stream flow of this module (hand pose frame sequence and string-plucking event log) will be stored synchronously into a local database to be further evaluated quantitatively.

Table 1 shows the quantitative performance of the hand posture capture module under four typical guzheng fingering techniques. The data shows that the single-finger plucking

TABLE 1. Accuracy evaluation of the hand posture capture module under different finger movements.

Finger type	Average error of fingertip spatial coordinates (mm)	Average error (degrees) in finger bending angle	Effective string-picking event detection rate (%)	Average system latency (ms)
Single-finger plucking (middle finger)	0.43 ± 0.11	2.1 ± 0.8	98.7	18.4
Big pinch (thumb + middle finger)	0.61 ± 0.18	3.4 ± 1.2	96.2	19.1
Quick flick (thumb)	0.52 ± 0.15	2.8 ± 1.0	95.5	19.8
Two hands, four fingers in a twirling motion	0.89 ± 0.27	4.5 ± 1.6	91.3	20.5

motion is the simplest, with the lowest fingertip coordinate error (0.43 mm) and the highest detection rate (98.7%). As the complexity of the fingering increases, such as the large plucking motion which requires simultaneous tracking of the coordinated movement of two fingers, the error slightly increases but remains within an acceptable range (average distance error between the thumb and middle finger tip is 0.61 mm). The fast plucking motion demands extremely high single-finger repetition speed from the thumb, resulting in a slight decrease in the detection rate to 95.5%, indicating that a very small number of rapid, light touches are not fully captured in high-frequency reciprocating movements. The four-finger tremolo with both hands places the greatest pressure on the system's tracking, with the average error rising to 0.89 mm, a detection rate of 91.3%, and a latency approaching the 20-millisecond limit. Overall, the module's tracking accuracy and real-time performance for the most commonly used fingering techniques (plucking, plucking, sweeping, and pinching) at the beginner to intermediate levels meet the basic requirements for teaching feedback [11,12].

B. MULTIMODAL FEEDBACK MECHANISM AND COGNITIVE LOAD CONTROL STRATEGY

Cognitive resources of learners in VR teaching settings are preoccupied concurrently with the sense of vision, hearing, vestibular, and proprioceptive information. Cognitive load can be increased by overload of feedback information or conflict between different modalities which in turn can decrease learning efficiency. Thus, the essence of designing this module does not lie in the attempt to superimpose the feedback channels but in the creation of a complex of multimodal feedback scheduling rules, which are to vary with the task difficulty [13]. The simplest form of information during guzheng performance is auditory feedback. The system has real-time synthesized sound technology which is a physical model. In contrast to the traditional sample playback, the synthesis model computes the string vibration mode in real-time according to the parameters of the string-plucking event (string speed, string position, and the angle of the string-plucking direction) as in module 3.1. In particular, the amplitude of the timbre-determined string vibrations depends exponentially on the speed of the string, and the ratio of high-frequency components in the timbre rises as the point of string-plucking draws towards the bridge. This synthesizer is provided as a synthesizer plug-in in the Unreal Engine 4.27 audio engine and has an output sampling rate of 48kHz and a latency of under 5 milliseconds. This makes sure that the change of timbre as perceived by learners in VR is greatly in line with the physical features of the real guzheng, hence maintaining the essence of the traditional guzheng teaching that is to correct through listening.

The visual feedback will be structured in an overlay mode. The bottom layer is a realistic simulation of the virtual guzheng and hands, giving a bare minimum of presence. The enhancement layer produces additional visual information in the vicinity of the hand model of the learner, dynamically triggered by the type of training task being performed. To illustrate, during finger placement exercise, the target string will show a gold light effect, with the fingertip that matches the proper shape of the plucking hand showing a semi-transparent blue guide ball. During rhythm training, a short (0.1-second) burst of a green pulse light effect will burst at the point of contact between the fingertip and the string when the plucking time of the learner is less than half of the standard metronome time; a yellow light effect will be shown when the deviation is between 50 and 150 milliseconds; and a red light effect will be shown when the deviation is greater. More importantly, based on preliminary pilot tests (N=10) correlating behavioral errors with the NASA-TLX scale, once three consecutive string-plucking instances are rated as invalid plucks (detected but unmet condition) or two consecutive rhythm deviations exceed 150 milliseconds, the system concludes that the current learning challenge likely exceeds the learner's immediate cognitive threshold. While these failures can occasionally stem from sensor occlusion, our pilot data indicated a 88% correlation between these specific error patterns and reported subjective overload. To prevent false interventions, the system requires these conditions to be met within a narrow 5-second window. It will momentarily mask all other visual information in the enhancement layer (leaving only the basic realistic layer) and a 2-second calming sound effect to encourage the learner to stop the action and breathe deeply. The intention of this strategy is to actively intervene and decrease temporarily too much cognitive load.

In the case of haptic feedback, given that no consumer-grade VR devices have haptic fingers with high-precision force feedback, this paper takes a different approach. Vibration amplitude and duration are mapped to the touch speed of a plucking event by SDK programming on the internal motor of the wireless handheld controller (VIVE controller). As the learners will not be touching the plucking paddle, this controller vibration is a cross-modal sensory substitution. While the effectiveness of this sensory substitution remains limited compared to real string resistance, a preliminary internal pilot (N=10) showed that learners using controller vibrations achieved 15% higher rhythmic accuracy compared to a 'no-haptic' VR condition. At an input speed of 0.3 to 1.5 m/s linearly mapped onto an output strength of 0 to 255 at a constant response duration of 50 milliseconds, learners can cross-modally associate the hand grip sensation with auditory timbre and visual light effects, partially compensating for the absence of a physical string-plucking channel. The parameters of the three feedback channels described above can be adjusted individually or turned off using the interface of the teacher to suit the requirements of the various stages of learning.

Table 2 clearly shows the non-linear relationship between feedback intensity and learning performance. When only the basic realistic layer was provided, learners had the lowest subjective cognitive load (5.2 points), but also the lowest fingering accuracy (68.3%), indicating insufficient practice efficiency due to a lack of guidance. Adding target highlighting and a guide ball increased the cognitive load to 6.8 points, but significantly improved fingering accuracy to 79.5%, and shortened task completion time by 18.2 seconds. Further adding rhythmic light effects further improved accuracy to 84.1%, but the cognitive load increased to 7.9 points, and learners' gaze switching frequency reached 31.7 times per minute, meaning they frequently shifted their gaze between the strings, fingers, and light effects, easily leading to visual fatigue and distraction. However, when the adaptive cognitive load control mode was enabled, the system actively simplified visual feedback when it detected unstable learner performance, causing the cognitive load to drop to 6.1 points while maintaining a high accuracy of 82.3%, and the gaze switching frequency decreased to a more comfortable 19.6 times per minute. This result validates the effectiveness of performance-based dynamic feedback adjustment strategies in VR teaching environments.

C. TASK-DRIVEN TEACHING SEQUENCE AND QUANTITATIVE EVALUATION INDICATORS

To attain structured teaching, a sequence of practice based on tasks is implemented into the VR environment. This order cites the fundamental methods needed in amateur guzheng tests at levels one to three, and is separated into five progressive steps. The initial step is concerned with one-string positioning and simple fingering (hooking, plucking and smoothing), including eight practice units. Learners are required to do repeated plucking with a certain finger of one string with visual guidance, but no time pressure in each unit. The second phase presupposes an alternating practice on neighbor strings (smoking and plucking, hooking and plucking), six units, introducing the elements of rhythm and making learners play cross-string under the conditions of a metronome that plays 60 beats per minute. The third stage is about simple scales and short phrases (ascending and descending D-tone pentatonic scale), which is made up of four units, and the first unit stresses continuity and even dynamics. The fourth phase comprises plain musical passages with big and little plucks (like a passage in Little Cabbage), consisting of three units. The fifth phase is a full-minute test piece (the prelude to "Phoenix Song"), which entails a tempo of 100 beats/minute and an 85-percent accuracy rate [14,15].

The practice units are packaged as task capsules with three sub-modules: a demonstration module, a practice module and a playback evaluation module. The demonstration module uses a semi-transparent virtual hand that displays three movements at a standard speed and the camera automatically centers on the important parts of the movements (like the point of touch on the finger and the string). When the practice module is on the system records real time performance data of the learner and compares it to a preset standard model. The standard model is not constructed based on one performer but rather created by averaging the data of five professional teachers of guzhenk who play the same etude by using Dynamic Time Warping (DTW). The quantitative variables to be compared are mostly the average distance between the fingertip and the center of the target string (measured in millimeter); the absolute value of the time difference between the plucking moment and the identical moment in the standard model (measured in milliseconds); the coefficient of variation of the interval time between two consecutive pluckings (measured in percent); and the root mean square error between the angle of the finger bending The replay assessment module provides a quantitative report automatically, which is shown on a floating panel in the VR headset, after the practice is completed. The report does not give a single, general score but displays the status of the four indicators in a radar chart (red areas indicate the level of performance of the learner and gray dashed lines indicate the passing grade). At the same time, the system gives textual feedback in response to errors, like "When plucking the strings with your middle finger, the tip of the finger is 3 millimeters out to the left; please attempt to move your wrist to the right in line with the breathing prompts of the metronome" or The rhythm is too slow; the current interval is 80 milliseconds longer than the standard; please adjust your wrist to the right in Each complete sequence of data produced in practice (timestamps, parameters of each plucking event) is uploaded to a cloud server where it is analyzed and intervened by the teacher.

IV. RESULTS AND DISCUSSION

A. THE IMPACT OF VR TEACHING ENVIRONMENT ON FINGER TECHNIQUE STANDARDIZATION AND RHYTHM STABILITY

To evaluate the actual teaching effectiveness of the designed VR teaching environment, this study recruited 60 college students with no prior experience learning the guzheng (Chinese zither). They were randomly divided into an

TABLE 2. Cognitive load and task performance of learners under different visual feedback intensities (N=30 beginners, task: complete 30 simple pieces of music)

Visual feedback mode	Subjective cognitive load score (1-10 points)	Average task completion time (seconds)	Fingering accuracy (%)	Visual fixation switching frequency (times/minute)
Basic realistic layer (no enhancement)	5.2 ± 1.1	187.4	68.3	12.1
Enhancement layer (target highlight + guide ball)	6.8 ± 1.4	169.2	79.5	23.4
Enhancement layer + rhythmic lighting effects	7.9 ± 1.6	158.6	84.1	31.7
Adaptive adjustment (cognitive load control)	6.1 ± 1.2	163.5	82.3	19.6

experimental group (VR teaching group, 30 participants) and a control group (traditional one-on-one teaching group, 30 participants). Informed consent was obtained from all participants. Both groups were identical in teaching duration (twice a week, 45 minutes each time, for 8 weeks), teaching content (the practice sequence of the five stages mentioned above), and the final assessment piece. The only difference was that the experimental group used the VR environment designed in this study for practice, while the control group learned in a physical classroom with a senior guzheng instructor with over fifteen years of professional teaching experience using a real guzheng. To ensure a fair comparison, the control group followed a strictly standardized manual curriculum that mirrored the exact five-stage practice sequence, task difficulty, and time allocation programmed into the VR system, thereby minimizing confounding variables related to instructional structure. All participants underwent a pre-test before the intervention, which involved playing a very simple single-note melody on a real guzheng to confirm that there was no significant difference in baseline performance between the two groups (independent samples t-test, $p > 0.05$). Eight weeks later, both groups performed the final assessment piece on the same real guzheng in the same physical classroom, and were scored by three experts who were not involved in the teaching, following a blind review process. It should be noted that the experimental group faced a 'transfer-of-learning' challenge during this assessment; while their VR training relied on controller vibrations, the final test required adapting to the physical string resistance and tension of a real instrument. The system logs and post-test interviews suggest that while finger positioning transferred effectively, the experimental group required a brief 5-minute familiarization period with the real guzheng to calibrate their tactile force application. The scoring sheet includes three dimensions: fingering accuracy (out of 100), rhythmic stability (root mean square of time deviation from the standard beat, in milliseconds), and tone purity (out of 100).

According to Table 3, the experimental group performed significantly better in core dimensions of core finger technique standardization and rhythmic stability compared to the control group. In terms of finger technique standardization, the experimental group had a mean of 86.7 points and the control group had a mean of 64.3 points with a difference of 22.4 points with a Cohen d of 2.34 which is a very strong effect. This shows how accurate hand posture capture and real-time visual instructions (e.g. a guide ball and target highlighting) in the VR environment effectively enable beginners to form the right muscle memory without the solidification of incorrect hand postures that can be formed in conventional learning because the teacher cannot constantly monitor learning. The low score (64.3 points) and wide standard deviation (11.5 points) of the control group also indicate that individual differences are easily exaggerated in conventional teaching and some students might lag behind due to the inability to be corrected as quickly as possible.

The differences were even more pronounced in terms of rhythmic stability . The root mean square of the rhythmic deviation of the experimental group was merely 32.6 milliseconds or that the deviation of the standard model was in an extremely small range in most of the plucking moments (nearly inaudible to the human ear). Conversely, the difference in the control group was up to 87.0 milliseconds, which is in the realm of distinctly perceivable rushing, or slowing. The effect size d Cohens was equal to -3.17, a very large absolute value. This outcome could be explained by the fact that the metronome synchronization that is implemented in the VR environment is very precise and the feedback of rhythmic light effects (green pulse light effect offers immediate rewards to a correct strike of the string). Traditional teaching will usually involve teachers counting verbally or using external metronomes, whereas students with traditional teaching methods will not have the immediate sensory experience of how small the gap between plucking and the beat is. This difference in

Time is converted to visual experience by the VR environment, which significantly enhances rhythm training efficiency and accuracy. Nonetheless, when it came to timbre purity the two groups did not differ significantly ($p=0.08$), with the experimental group scoring marginally higher by 7.2 points, but the effect size was less (0.62). This is probably due to the fact that the quality of the tone is very much dependent on the angle of plucking, depth and the texture of the fingernail on the string. Although the existing VR devices are capable of creating relatively lifelike tones using physical models, the haptic feedback mastered by learners by using controller vibrations cannot fully replicate the nuanced haptic experience of a physical pick touching a string. This suggests that more advanced finger-type force feedback devices need to be integrated in the future.

TABLE 3. Comparison of assessment scores between the experimental group and the control group after eight weeks of teaching intervention (N=60)

Evaluation Dimensions	Experimental group (VR teaching group, n=30)	Control group (traditional teaching group, n=30)	Mean difference	Effect size (Cohen's d)
Finger placement accuracy score (points)	86.7 ± 7.2	64.3 ± 11.5	22.4	2.34
Root mean square of rhythm offset (milliseconds)	32.6 ± 9.8	87.0 ± 21.3	-54.4	-3.17
Tone purity rating (points)	78.4 ± 10.1	71.2 ± 12.6	7.2	0.62

B. ANALYSIS OF LEARNERS' BEHAVIORAL PATTERNS AND COGNITIVE STRATEGIES IN VR ENVIRONMENTS

Along with the final assessment scores, the process data, produced by the learners in the experimental group within the VR environment were also analyzed to unveil the behavior patterns and mental strategies of the learners. System logs provided the data, which included the total time that a learner was interacting with the system during each practice session, the dwell time in each feedback mode, error correction time and the frequency of pauses. A K-means cluster analysis (with K=2 determined by the elbow method and Silhouette coefficient) using Euclidean distance metrics for variables including gaze duration and error-correction latency was performed on more than 240 hours of practice logs. This revealed two overall patterns of behavior. Statistical analysis (independent samples t-test) further confirmed that the 'auditory-first' group showed sig nificantly higher resilience ($p < 0.05$) in complex stages compared to the 'visual-dependent' group. The former pattern (63 percent of the sample) was termed as visual-dependent, which involves a pattern of scanning visual guidance elements within the enhancement layer (guide balls and rhythmic light effects) in the course of practice and the gaze heatmap was concentrated between the center of the strings and the hand model. Students in this pattern showed a temporary drop in performance as soon as visual guidance (entry of adaptive adjustment mode) was removed, but regained performance fairly quickly afterwards. The second mode (37 percent of the sample) was identified as auditory-first, which was more dominated by the distribution of the gaze points on the left side of the virtual guzheng (non-direct playing field) and less reliant on the effect of rhythmic lighting. Even with auditory feedback (synthesized sounds and metronome) its rhythmic accuracy is capable of a high level.

Further breakdown showed that initially, the visually-dependent learners would develop faster (with fingering accuracy, improving 35% faster during the first four weeks) but when reaching the fifth stage (complex musical phrases), they would take an average of 42 seconds longer than auditory-first learners to correct their error. This indicates that excessive dependence on visual feedback can result in a crutching effect in which learners become unable to change their focus off the external visual feedback to internal kinesthetic and auditory feedback during complex tasks. On the other hand, although the auditory-first learners had a slow start, they were more fluent and resilient in sustained performance when they formed stable internal auditory representations. According to this observation, a visual feedback decay mechanism should be included in VR teaching systems in the future. This implies that as learners gain greater proficiency with a practice unit (more than 90% three consecutive attempts) contrast and brightness of the visual enhancement layer should be progressively diminished, compelling learners to depend more on the auditory and proprioceptive senses, and facilitating the process of skills being reclassified as internal control as opposed to external control. Moreover, the data indicate that during practice, learners took an average of 6.8 minutes before having some form of a behavioral pause (i.e., no effective strumming over a period of over 5 seconds with a head turn of less than 5 degrees). The further examination of these pauses showed that 62% of these pauses took place in 5 seconds after negative feedback (a red light effect or a large deviation warning). This may be construed as short-term cognitive reflection or emotional control by learners. Assuming that the system is capable of proactively playing a short and encouraging voice message (like: Your deviation was a bit large just now, it is okay, slow down and try again) once signaled the presence of such a pause after negative feedback, it was capable of reducing the recovery time of the learner by an average of 4.3 seconds and enhancing the likelihood of again correctly strumming by 18%. This finding reveals the significance of considering social-emotional support in VR instructional settings to prevent cold and quantitative feedback that may diminish the learning motivation.

C. ANALYSIS OF THE LIMITATIONS AND TECHNOLOGICAL BOTTLENECKS OF THE TEACHING ENVIRONMENT

Despite the fact that the VR guzheng teaching environment developed in this research has proven to be effective based on various measures, there are still a number of critical limitations and technical bottlenecks to the effective implementation and promotion of this environment into practice. To begin with, there are the problems of expense of hardware and comfort of wear. This system is based on a combination of HTC VIVE Pro Eye and Leap Motion 2.0, and the whole system costs more than 15,000 RMB, which is far beyond the economic ability of average individual learners. More to the point, due to the weight of the head-mounted display (around 750 grams) and cable limitations, together with extended wear (more than 30 minutes), the physiological load can be a huge burden during the learning of the instrument, which may be compounded by neck-strain and pressure on the face. In the research, about 15 percent of the respondents said that they stopped their practice in the second half because of headache.

Secondly, there exist physical constraints in terms of tracking distance and motion grip. Leap Motion's field of view is 150 degrees. At very high speeds, or when the sensor is exceeded by the movement of the learner hand (e.g. when playing high and low notes with both hands at the same time, one of the hands can approach the edges of the field of view), tracking will either lose the track or introduce a visible jittering effect. Even in complex movements, like four-finger tremolo movements, the mutual occlusion of the thumb and index finger may result in the finger ID swapping (that is, the model fails to recognize the index finger as the thumb), thereby causing incorrect feedback. Even though the study was keeping a detection rate of 91.3% with the optimization of the algorithm, false detection rate of almost 10 percent is unacceptable among the advanced learners in the quest to achieve the final performance. For the future, it is necessary to consider hybrid tracking solutions with multi-camera arrays or data gloves.

Thirdly, there are limitations in the realism and expressiveness of the timbre physical model. Although the parameters of current real-time synthesizers can be used to alter the timbre, depending on the speed and position of the touch of the strings, these models only take into account the vibration of a perfect, homogeneous string. There is also the complicated interaction of timbre of a real guzheng of the acoustic response of the resonating cavity in the instrument, the cross-induction of the strings, and slight variations in the fingernail material and the angle of touch of the strings by the player. The timbre purity in the experimental group was not significantly greater than it was in the control group in the study, which indirectly proves this limitation. A more realistic acoustic model of finite elements may be more realistic but its calculation is extremely expensive (it needs a calculation time of less than 10 milliseconds per frame), compared to the real-time capabilities of existing VR devices. This is a typical dilemma of balancing accuracy and real-time performance.

Lastly, there is the danger of single teaching evaluation system. At the moment, VR systems do not even measure such quantifiable signs (positional error and time deviation) and do not in any way consider the more important artistic aspects of guzheng performance, musical expressiveness, emotional expression, and breath management of musical phrases. The use of it in the long term can result in learners unconsciously seeking out mechanical perfection at the cost of the emotional flow of the music. Thus, the VR setting must be well placed as a skills training tool and not as the subject of artistic critique, and its future usage in teaching must be accompanied by the traditional humanistic teacher guidance.

V. CONCLUSION

This study focuses on an immersive teaching environment for guzheng performance based on virtual reality technology, completing a full closed loop from technical design and module construction to empirical evaluation. First, a VR system was designed and implemented, including a high-precision hand posture capture module, an adaptive multimodal feedback mechanism, and a task-driven teaching sequence. The technical parameters and implementation logic of each module are detailed. Second, through an eight-week controlled experiment, empirical results shown in three key data tables demonstrate that the VR environment has a significant immediate effect on improving the fingering accuracy and rhythmic stability of beginners. System optimization strategies such as "visual feedback attenuation" and "emotional computing support" are proposed. Finally, the technical bottlenecks of the current system in terms of hardware cost, tracking occlusion, timbre realism, and the simplification of the evaluation system are objectively analyzed.

The main contributions of this study are: First, it provides a reproducible and quantifiable technical solution and empirical data foundation for the digital transformation of ethnic musical instrument teaching; second, it verifies the effectiveness of performance-based adaptive adjustment strategies for cognitive load in fine motor skills learning scenarios; and third, it reveals the correlation between different behavioral patterns and learning outcomes in VR teaching environments, providing a basis for personalized learning path design. Future work will focus on lightweight hardware integration (such as standalone VR headset solutions), introducing multi-view hand tracking networks to address occlusion issues, and exploring the integration of generative AI technology to provide learners with more musically expressive demonstrations and evaluations.

AUTHOR CONTRIBUTIONS

Miao Wang designed, collected and analyzed the data, and drafted the manuscript. Miao Wang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Miao Wang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] P. Phakamach, P. Senarith, and S. Wachirawongpaisarn, "The Metaverse in Education: The Future of Immersive Teaching & Learning," RICE Journal of Creative Entrepreneurship and Management, vol. 3, no. 2, pp. 75-88, 2022, doi: 10.14456/rjcm.2022.12.

[2] T. Roche, E. Wilson, and E. Goode, "Immersive learning in a block teaching model: A case study of academic reform through principles, policies and practice," Journal of university teaching and learning practice, vol. 21, no. 2, pp. 1-29, 2024, doi: 10.53761/1.21.2.12.

[3] H. Khamis, M. Khairudin, H. Jantan A, et al., "Systematic review of adapting immersive technology in enhancing teaching and learning for students in higher education," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 61, no. 3, pp. 130-145, 2024, doi: 10.37934/araset.63.1.130145.

[4] L. Lowell V and W. Yan, "Applying systems thinking for designing immersive virtual reality learning experiences in education," TechTrends, vol. 68, no. 1, pp. 149-160, 2024, doi: 10.1007/s11528-023-00922-1.

[5] A. Bizami N, Z. Tasir, and N. Kew S, "Innovative pedagogical principles and technological tools capabilities for immersive blended learning: a systematic literature review," Education and information technologies, vol. 28, no. 2, pp. 1373-1425, 2023, doi: 10.1007/s10639-022-11243-w.

[6] D. T. K. Ng, E. H. L. Ng, and S. K. W. Chu, "Engaging students in creative music making with musical instrument application in an online flipped classroom," Education and information Technologies, vol. 27, no. 1, pp. 45-64, 2022, doi: 10.1007/s10639-021-10568-2.

[7] M. Yin, "Music teachers' professionalism: Realizing intercultural competence in guzheng education when using a MOOC," Education and information technologies, vol. 28, no. 10, pp. 13823-13839, 2023, doi: 10.1007/s10639-023-11710-y.

[8] Y. Huang, S. Chuangprakhon, and P. Santaveesuk, "Preservation and transmission of Shaanxi Guzheng musical instruments: Challenges and strategies for cultural sustainability," International Research Journal of Multidisciplinary Scope, vol. 5, no. 04, pp. 147-158, 2024, doi: 10.47857/irjms.2024.05i04.01265.

[9] Z. N. Khlaif, A. Mousa, and M. Sanmugam, "Immersive extended reality (XR) technology in engineering education: Opportunities and challenges," Technology, Knowledge and Learning, vol. 29, no. 2, pp. 803-826, 2024, doi: 10.1007/s10758-023-09719-w.

[10] S. M. Lee and J. G. Wu, "Preparing teachers for the future: Microteaching in the immersive VR environment," ReCALL, vol. 36, no. 3, pp. 251-269, 2024, doi: 10.1017/S0958344024000089.

[11] I. Zinkiv and J. Ren, "CHINESE ZITHER SE IN ANCIENT AND TRADITIONAL MUSICAL AND INSTRUMENTAL CULTURES OF TURKISH, MONGOLIAN AND TUNGUSKA-MANCHU ETHNIC GROUPS," Yegah Müzikoloji Dergisi, vol. 7, no. 4, pp. 510-533, 2024, doi: 10.51576/ymd.1543588.

[12] K. Mehta, K. M. Nayak, R. A. Qureshi, et al., "Technology acceptance and SAMR in metaverse education: professors' willingness to engage in immersive classes in higher education," Journal of Applied Research in Higher Education, vol. 18, no. 3, pp. 830-849, 2026, doi: 10.1108/JARHE-08-2024-0430.

[13] T. Vaizman, "Teaching musical instruments during COVID-19: teachers assess struggles, relations with students, and leveraging," Music Education Research, vol. 24, no. 2, pp. 152-165, 2022, doi: 10.1080/14613808.2022.2053512.

[14] A. Förster, "Accessible digital musical instruments in special educational needs schools–design considerations based on 16 qualitative interviews with music teachers," International Journal of Human–Computer Interaction, vol. 39, no. 4, pp. 863-873, 2023, doi: 10.1080/10447318.2022.2049141.

[15] J. Stijnen, L. Nijs, and P. Van Petegem, "Instrument teachers' practices, beliefs, and barriers regarding musical creativity: Exploring the creative process of interpretation," International Journal of Music Education, vol. 42, no. 3, pp. 425-441, 2024, doi: 10.1177/02557614231175777.