International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6145-6158

https://doi.org/10.70517/ijhsa463528

Optimization of personalized piano technique training based on the Monte Carlo algorithm: Empowering innovation in college music education

Luyao Liu[1,*] and Weiyu Zhu[2]
[1] School of Music, Shandong University of Art, Jinan, Shandong, 250014, China
[2] School of Music Education, Sichuan Conservatory of Music, Chengdu, Sichuan, 610021, China
Corresponding authors: (e-mail: skdwang0314@163.com).

Abstract Music education in colleges and universities is transforming towards digitalization and intelligence, and piano teaching, as the core curriculum of music education, faces the demand for technological innovation. Traditional piano teaching relies on teachers' subjective judgment in terms of hand shape correction and technique training, and lacks objective quantitative standards. This study constructs a piano technique training optimization system based on computer vision technology, aiming to improve the scientificity and accuracy of piano teaching in colleges and universities. Firstly, the piano string vibration equation model is established to analyze the acoustic features such as pitch and overtones; secondly, MEMS inertial sensors and infrared detection rods are used to collect the playing gesture data, and the fusion of fixed posture is realized by the IU-EKF algorithm; then, the VGG-16 deep learning model is used to extract the statistical features in time domain, spatial characteristics, finger coupling features and auxiliary features, and to realize the recognition of hand gestures; and finally, the Magic King is performed in different playing versions. The speed-strength visualization analysis is carried out for different playing versions. The results show that the average values of finger angle measurements are 147.92°, 136.03° and 117.15°, respectively, with a maximum error of only 2.73%; the maximum angular difference between the three paths of finger movement is only 2.6 degrees; the velocity calibration method effectively matches the finger sliding and rebounding velocities within the 95% confidence interval; and the predicted values of the music skill evaluation model are highly consistent with the actual values. The system proposed in this study can accurately identify piano playing gestures, provide objective quantitative indexes for piano technique training, and put forward optimization measures from three dimensions: fingering practice, technical difficulty attack, and experience learning, which is of great significance to promote the modernization of piano education in colleges and universities.

Index Terms Computer vision, Piano technique training, Gesture recognition, Deep learning, Performance analysis, Music education

I. Introduction

Music and computer are two completely different fields and have essential differences, however, with the continuous development of the times, this difference is gradually narrowing, and they form a complementary relationship [1]. Entering the 21st century, computer technology is penetrating into all walks of life with a rapid speed, and the informatization and digitization technology it brings has covered our work and life, also in the field of music [2]-[4]. In fact, the use of computer music technology has been popularized in composition, music theory teaching, music production, recording and other musical arts, and after years of updating, the development and application of this technology has become more mature [5].

In recent years, due to the continuous updating of computer technology, coupled with the deepening call for piano teaching reform, the use of computer technology in piano teaching has been realized under the impetus of this wave [6]. Research and treatises on computer-assisted piano teaching have also appeared in large numbers, such as “the use of multimedia teaching methods in piano teaching”, “analysis of piano rhythm teaching in the realization of SONAR8 software”, “the new impetus for the reform of piano music education in China - the new model of digital piano music teaching”, all of which have started from the question of how to teach piano music with the help of Computer music in piano teaching rhythm training, the cultivation of students' interest, teaching forms, teaching methods of reform and other multiple perspectives to put forward their own ideas and research conclusions [7]-[9]. This gives the overall level of piano teaching, there is a greater improvement. Accompanied by the rapid development and updating of computer technology, the sequencing software has its powerful functions, humanized and simple operation interface management, and has gradually become a teaching means to improve the quality

of teaching [10]-[12]. With diversified computer vision technology, it also plays an increasingly important role and efficacy in the actual piano teaching [13].

Music education plays an irreplaceable role in cultivating students' comprehensive quality and improving their artistic cultivation, while piano teaching, as an important part of the music education system, has a direct impact on the quality of teaching and the effect of cultivating students' musical literacy. The traditional piano teaching mode mainly relies on teachers' experience judgment and subjective evaluation, and lacks objective quantitative standards in technical aspects such as hand shape correction, key touch strength, rhythm control, etc. The teaching effect is often uneven due to the difference in teachers' level. In recent years, the rapid development of computer vision, deep learning and other artificial intelligence technologies has provided new ideas and methods to solve the technical problems in piano teaching. The real-time capture and analysis of the player's hand movements through computer vision technology can accurately identify key parameters such as finger position, key touch angle, and strength change, providing a scientific and quantitative basis for piano technique training.

Starting from the physical characteristics of piano music, this study first establishes the piano string vibration equation, reveals the physical mechanism of piano sound generation through theoretical derivation and mathematical modeling, and analyzes the formation principles of acoustic features such as pitch, overtones, and dissonant overtones, so as to lay a theoretical foundation for the subsequent technical analysis. On this basis, a multimodal data acquisition system is constructed by using micro-inertial sensors and infrared detection technology to obtain the attitude information of each part of the hand during the playing process, and the data fusion and attitude estimation are realized by iteratively updating the extended Kalman filter algorithm. Then we extract multi-dimensional features such as time domain statistical features, spatial characteristics features, inter-finger coupling features, etc., and use VGG-16 deep learning model for gesture recognition and classification. Finally, different performance versions of the classic piano work “The Magic King” are selected for case study, and the differences in the performance styles of different performers are compared through the tempo-strength visualization technique, so as to refine the optimization strategy of piano technique training.

II. Characterization of piano music

II. A. Piano string vibration equation and its acoustics

II. A. 1) Theoretical derivation of the equations of vibration of piano strings

The articulation process of a modern piano involves the player touching the keys and then striking the strings by means of a mechanically conducted felt hammer, so that the vibrations of the strings are excited by the felt mallet. The mallets themselves are attached to the keys of the piano, and when the keys are pressed, the mallets strike the strings and vibrate the strings to make the piano sound. Under normal playing conditions, the force exerted by the mallets on the strings is limited to the small area where the mallets come into contact with the strings, and the strike is very brief.

Suppose that the string is l long and is fastened to the sounding board at both ends by tuning pegs, and is now struck by a cosine shaped convex mallet of width 2δ, the interval over which the mallet strikes is $(x_0-\delta \le x \le x_0+\delta)$. Let the point of hammering be at $x=x_0$, the initial velocity of the string motion be v_0, the acceleration of the string vibration be a, and the time of string vibration be t. Since it is a cosine-type convex mallet, at the moment of striking, the point $x=x_0$ obtains the maximum velocity, while at the points $x=x_0-\delta$ and $x=x_0+\delta$ the velocity is 0. After that the string starts to vibrate freely, and the vibration of the string can be reduced to a fixed-solution problem:

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2}=a^2\dfrac{\partial^2 u}{\partial x^2},(0<x<l,t>0) \\ u|_{x=0}=u|_{x=l}=0 \\ u|_{t=0}=\varphi(x)=0 \\ \dfrac{\partial u}{\partial t}|_{t=0}=\begin{cases} v_0\dfrac{\cos\dfrac{x-x_0}{2\delta}\pi}{2\delta}, & (x_0-\delta\le x\le x_0+\delta) \\ 0, & (0\le x\le x_0-\delta, x_0+\delta\le x\le l) \end{cases} \end{cases} \tag{1}$$

According to the boundary conditions, the eigenfunction is known to be $\sin\dfrac{n\pi x}{l}$, which can be expressed according to the general formula for string vibration:

$$u(x,t)=\sum_{n=1}^{\infty}\left(C_n\cos\frac{n\pi a}{l}t+D_n\sin\frac{n\pi a}{l}t\right)\sin\frac{n\pi}{l}x \tag{2}$$

Bring Eq. (2) into the initial conditions to determine the coefficients C_n and D_n:

$$\varphi(x)=\sum_{n=1}^{\infty}C_n\cdot\sin\frac{n\pi}{l}x=0\quad 0<x<l \tag{3}$$

$$\begin{aligned}\psi(x)&=\sum_{n=1}^{\infty}D_n\frac{n\pi a}{l}\cdot\sin\frac{n\pi}{l}x\\&=\begin{cases}v_0\dfrac{\cos\dfrac{x-x_0}{2\delta}\pi}{2\delta}, & (x_0-\delta<x<x_0+\delta)\\0, & (0<x<x_0-\delta,x_0+\delta<x<l)\end{cases}\end{aligned} \tag{4}$$

The left side of Eqs. (3) and (4) is the Fourier series of that function on the right, from which the coefficients can be derived as follows:

$$C_n=\frac{2}{l}\int_0^l\psi(x)\sin\frac{n\pi x}{l}dx=0 \tag{5}$$

$$\begin{aligned}D_n&=\frac{2}{n\pi a}\int_0^l\psi(x)\sin\frac{n\pi x}{l}dx\\&=\frac{8v_0\delta}{n^2\pi^2a}\frac{1}{1-\dfrac{4\delta^2n^2}{l^2}}\sin\frac{n\pi x_0}{l}\cos\frac{n\pi\delta}{l}\end{aligned} \tag{6}$$

So the equation for the vibration of a piano string is:

$$\begin{aligned}u(x,t)=&\frac{8v_0\delta}{\pi^2a}\sum_{n=1}^{\infty}\frac{1}{n}\frac{1}{1-\dfrac{4\delta^2n^2}{l^2}}\sin\frac{n\pi x_0}{l}\\&\cdot\cos\frac{n\pi\delta}{l}\sin\frac{n\pi at}{l}\sin\frac{n\pi x}{l}\end{aligned} \tag{7}$$

II. A. 2) On the piano string vibration equation and its piano acoustics

Audio refers to other sounds in addition to human speech and music, including the sound of the natural environment, animal sounds, the sound of machines and tools, and a variety of sounds emitted by human action.

Sound is roughly including amplifiers, peripherals (including pressure limiters, effects, equalizers, VCDs, DVDs, etc.), speakers (speakers, horns), mixing consoles, microphones, display devices and so on add up to a set. Among them, the speaker is the sound output device, speakers, subwoofer and so on. A speaker includes high, low and center speakers, three but not necessarily three.

For the acoustics of a piano, many researchers have studied and come to conclusions.

(1) Pitch

The pitch heard by the human ear is the pitch corresponding to the fundamental frequency of the vibration of the piano strings, i.e., the fundamental tone determines the corresponding pitch of the keys [14]. The human ear cannot hear all of the octaves, but experiments have shown that in certain environments the human ear can distinguish some of the octaves. In fact, the octaves that can be heard by the human ear have a major influence on the timbre of the sound. Various musical instruments have their own special sound color, because of the sound contained in the octave and the octave of the intensity of the difference caused by the sound.

(2) Overtones

From the equation, it is seen that the piano string vibration is a superposition of all standing waves and n th harmonic in amplitude. The overtones of the sine and cosine functions decay proportionally to $1/n$, while the overtones of the plucked string vibration decay proportionally to $1/n^2$. It is because the piano string vibration contains a wealth of overtones, its sound is full and beautiful.

(3) Dissonant overtones

In the manufacture of actual pianos, the strike point $x = x_0$ is chosen at the 1/8 position of the string length, because the position where the mallet strikes the string is not a point, but an area of 2δ. In the treble, due to the shorter string length, the choice of the 1/8 position effectively suppresses the dissonant overtones in the 1/7th and 1/9th positions.

II. B. Characterization of piano music

II. B. 1) Preprocessing note sequences

Before the automatic labeling of chord fingerings, the note sequences are preprocessed, and the main preprocessing process is as follows.

First, the chordal music played on the piano is processed to obtain the spectrum, which is adaptively adjusted using dynamic range compression due to the large dynamic differences in amplitude of the music signal [15]. Secondly, the self gain adjusted linear spectrum is converted into Chroma spectrum, compared with the traditional method, there is no need to normalize the data when extracting the Chroma, so that the original linear spectral characteristics can be maintained, and the dimensionality of the output can also be reduced. The amplitude of the Chroma spectrum is compressed in the logarithmic compression formula, the expression of which is shown in Equation (8):

$$Z = \lg\left(1 + C \cdot Chroma\right) \tag{8}$$

where, C represents the degree of compression. After this step, each note will have 1 amplitude compressed Chroma eigen coefficient.

This is then obtained by normalizing the variance through the PCA space. At this stage, unsupervised learning is performed on the feature matrix after PCA dimensionality reduction using maximum pooling and averaging to obtain the feature vectors of the “bag of words”.

Finally, after the previous steps, the PCA whitened Chroma features are obtained, which are manually designed based on the existing knowledge, so in this step, a noise reduction encoder is used for processing. The automatic coding with noise suppression is a typical bottleneck model with the hidden layer as the actual output. During the training process, the optimal parameter weight matrix W and the deviation vector b are shown in Equation (9):

$$L = \min_{W,b} X_c - X_f + \lambda \left\| W \right\|_F^2 \tag{9}$$

where, L represents the noise suppression optimization objective parameters. X_c represents the whitened input layer. X_f represents the final output layer. λ represents the regular term parameters. $\left\| W \right\|_F^2$ represents the number of paradigms. After completing the training, the feature matrices of the initial points are learned to obtain the optimal parameter weight matrices and offset vectors. The basic features of chord fingerings are then obtained after processing with the noise autoencoder pattern of the optimal parameters and used as subsequent inputs.

II. B. 2) Allocation of labeled areas

After the above preprocessing, the piano playing chords are placed in a three-dimensional space, and the labeling in the sequence table is regarded as an element in the space. Assuming that the sequence labeled on any chord is at rest, the expression for the chord scale function is shown in Equation (10):

$$v = \sum_{i=1}^{m} X_i \tag{10}$$

where X_i represents the localization of the annotation and is at the highest position in the sequence, occupying the main part of the chord. The annotation layout method based on region partitioning is used to divide the annotation region into 8 regions, and the nodes covered by each graph are evenly distributed. After the partitioning process, the annotation layout is simplified and the same type of annotation will appear in the annotation region. For this reason, the same type of annotation is rearranged using sorting to avoid collision completely. In the arrangement, the annotation definition point is adopted as the index value of each region in order to form the corresponding annotation body. The processing in this stage is mainly to analyze the basic characteristics of the frequency spectrum of the performance, and complete the normalization process under the condition of certain frequency and unchanged sequence. On this basis, the received sound signal is captured, and the chord changes are utilized to speculate the nodes of the performance, adjusted to the corresponding amplitude to reduce the generation of errors, and the note sequences are preprocessed through the automatic editing of the instrument, with the flow shown in Figure 1.

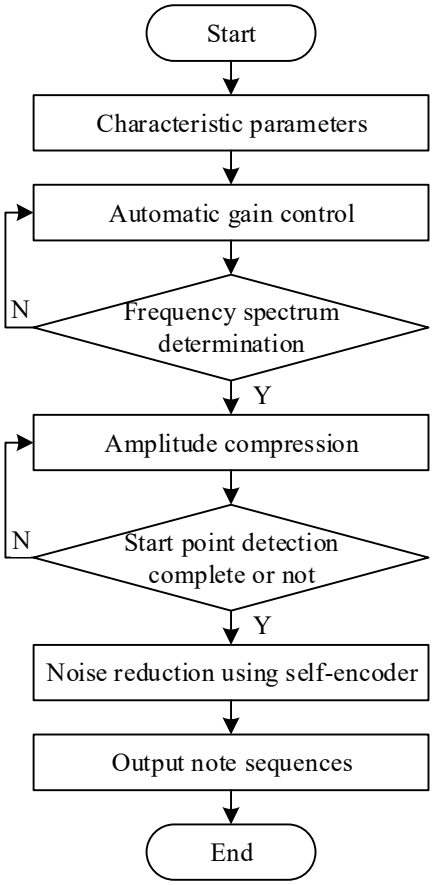


Figure 1: Note sequence pretreatment process

Following the steps in Fig. 1 to reduce the length of the chord, fusing the repeated notes, and detecting each note position labeled to form a dynamic matrix of note sequences, the expression is shown in Eq. (11):

$$A = \lg\left(1 + C' \cdot Chroma\right) \tag{11}$$

where, A represents the matrix. C' represents the reduction parameter.

Due to the preponderance of influences in processing the sequence, noise is occasionally generated, so an automatic editor that can reduce noise is used to recognize the chord fingerings in order to filter out the noisy noise and adjust the music to a normal tune. Assuming that the notes in the spectral signal sequence detected by the instrument correspond to the fingerings, the relational equation between the two is shown in Equation (12):

$$y(k) = \left[12\log_2\left(\frac{k}{N \cdot f_r} \cdot f_b\right)\right] \tag{12}$$

where, $y(k)$ represents the mapping function. k represents the coefficients. f_r represents the spectrum. f_b represents the noise function. N represents the weights to be connected.

The dispersed set of notes is distributed according to the size of the frequency, after which the music signals will automatically find their own labeling positions according to the arrangement of the notes, reducing the vacancies on the chords, and the remaining signals will fill in the vacant labeling positions, providing a reference for the subsequent automatic labeling of the fingering.

II. B. 3) Converting note sequences

The pre-processed labeled fingerings do not have the problem of excessive distance between fingers, nor do they have the error of crossing fingers in the process of playing. Establishing the fingering labeling model under this condition corrects the problem of too slow fingering change on the one hand, and improves the overall efficiency of playing on the other hand, which makes the joining of tones more natural and rapid, including the elevation and

modulation of the music articulated more smoothly, and increases the probability of judgment from other perspectives, the principle model is shown in Fig. 2.

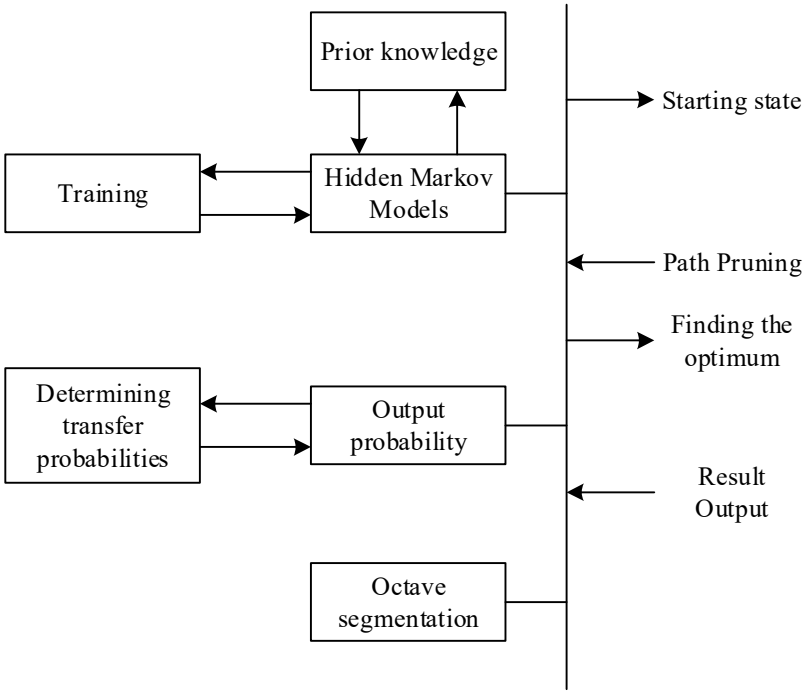


Figure 2: The construction of the index model

However, in the modeling process, the lifting and shifting between the fingers will affect the rhythmic changes and lead to errors in the chords. In order to improve the accuracy of the fingering labeling and obtain the corresponding information, the transfer probability of the note sequence is obtained by using the law of fingering transition under the premise that each parameter does not affect the calculation, as shown in Equation (13):

$$P\left(\beta \subseteq K\right) = \det\left(M_{\beta}\right) \tag{13}$$

where, K represents the probabilistic model. M_{β} represents the music clip.

When the note sequences in the set are input into the model, the system will randomly appear 1 matching fingering labeling method, and then detect the correctness of the fingering labeling and automatically form the matrix. Each note element in the matrix represents a different piece of information, and independent note sequences work better for automatic fingering labeling than collective filtering.

Assuming that the elements in the matrix are represented by y_1, and the positions of the parameters are deduced using the positions of the notes in the sequence, the expression for the initial state of the notes is shown in Equation (14):

$$F = \begin{cases} 6 - index(y_1) & K = 0 \\ index(y_1) & K > 0 \end{cases} \tag{14}$$

The notes will jump to a specific route according to the instructions of the matrix, removing the interference of noise, then the range of parameters labeled in the model is shown in Equation (15):

$$\sigma_n(i) = S_{\max} = \left(i_t = i \mid i_1, \cdots, i_n, \cdots, \lambda\right) i = 1, 2, \cdots, n \tag{15}$$

where, $S_{\max}$ represents the maximum value of the parameter. The λ represents the feature vector. All note sequences in the range can be transformed, but the model selects the sequence that best meets the criteria, and the remaining other routes are discarded so that the chord maintains a steady state.

III. Piano playing gesture recognition methods

This chapter proposes a deep migration learning based piano playing gesture recognition technique [16], which utilizes multimodal features to enhance the piano playing gesture recognition. The specific process is as follows:

(1) Micro-inertial sensors and infrared detection rods are used to collect the piano playing gesture data and obtain the gesture data of different parts of the hand. And the gestures are estimated by the state space model, after which the IU-EKF algorithm is utilized to realize the fusion fixed gesture.

(2) In order to obtain valuable gesture data of the active segment of the piano playing hand, the detection data of the infrared detection rod is acquired through a fixed-width sliding window to obtain the multimodal features of the piano playing gesture.

(3) The extracted multimodal feature data features are preliminarily classified using the infrared detection rod, after which four kinds of gesture features, namely, time-domain statistical features, inter-finger coupling features, spatial characteristic eigenvalues, and auxiliary features, are inputted into the Extreme Learning Machine model (VGG-16), and the feature model is trained to realize the classification and recognition of the piano playing hand gestures.

III. A. Piano playing gesture posture estimation and stance fixation

(1) Piano playing gesture posture estimation based on state space modeling

In this paper, micro-electro-mechanical system (MEMS) inertial sensors are utilized to collect hand data of piano playing gesture posture. After the acquisition is realized, the piano playing gesture posture is estimated by the posture estimation model.

Design state space model with piano playing gesture posture. In this paper, the quaternion is selected to describe the gesture and the following parameters are utilized as the sensor system state quantities as shown in Equation (16):

$$x = \begin{bmatrix} q_e^T & v_e^T & b_{a,e}^T & b_{g,e}^T \end{bmatrix}^T \tag{16}$$

The unit quaternion of the gesture pose in Eq. (16) is $q_e = \begin{bmatrix} q_{0,e} & q_{1,e} & q_{2,e} & q_{3,e} \end{bmatrix}^T$. The lower carrier velocity vector is $v_e = \begin{bmatrix} v_{east,e} & v_{north,e} & v_{up,e} \end{bmatrix}^T$. The velocity components along the skyward, eastward, and northward directions in the navigation coordinate system are denoted by v_{up}, v_{east}, and v_{north}, respectively. The accelerometer offset is $b_{a,e} = \begin{bmatrix} b_{ax,e} & b_{ay,e} & b_{az,e} \end{bmatrix}^T$. The gyroscope drift is $b_{g,e} = \begin{bmatrix} b_{gx,e} & b_{gy,e} & b_{gz,e} \end{bmatrix}^T$. T is the transposition identity. In accordance with the quaternion principle, the relationship between the attitude quaternion and the carrier angular velocity vector w can be identified as shown in Equation (17):

$$\dot{q} = \frac{1}{2}\Omega(w)q_e = \frac{1}{2}\begin{bmatrix} 0 & -w_z & -w_y & -w_x \\ w_z & 0 & -w_x & w_y \\ w_y & w_x & 0 & -w_z \\ w_x & -w_y & w_z & 0 \end{bmatrix}\begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} \tag{17}$$

In Equation (17), the antisymmetric matrix of the carrier angular velocity vector is $\Omega(w)$, the elements of the antisymmetric matrix of the carrier angular velocity vector are denoted by w_x, w_y, and w_z, the elements of the unit-quadratic post-transpose matrix of the gesture pose are denoted by q_0, q_1, q_2, and q_3, and the gyroscope output is given by $w = \tilde{w} - b_g - \eta_g$, $\tilde{w}$, and the gyroscope measurement noise is denoted by η_g.

The Jetlink inertial guide specific force equation is expressed through Eq. (18):

$$\dot{v} = R_b^a f^b + G_0 \tag{18}$$

In Equation (18), the rotation matrix from the piano playing gesture coordinate system to the navigation coordinate system can be described by the unit quaternion R_b^a, the scaling after compensating for the offset in the piano playing gesture coordinate system is f_b, and the gravitational acceleration vector is denoted by G_0. As shown in Eq. (19) and Eq. (20):

$$R_b^a = 2\begin{bmatrix} 0.5 - q_1^2 - q_2^2 & q_1q_2 - q_0q_3 & q_1q_3 + q_0q_2 \\ q_1q_2 + q_0q_3 & 0.5 - q_0^2 - q_3^2 & q_2q_3 - q_0q_1 \\ q_1q_3 - q_0q_2 & q_2q_3 + q_0q_1 & 0.5 - q_1^2 - q_2^2 \end{bmatrix} \tag{19}$$

$$R_b^a = \begin{bmatrix} \cos\varphi\cos\psi + \sin\varphi\sin\psi\sin\theta & \sin\psi\cos\theta\sin\varphi\cos\psi \\ & -\cos\psi\sin\varphi\sin\theta \\ -\cos\varphi\sin\psi & \cos\varphi\cos\psi\cos\theta \\ +\sin\varphi\cos\psi\sin\theta & -\sin\varphi\cos\psi - \cos\varphi\sin\psi\sin\theta \\ \sin\varphi\cos\theta & \sin\theta\cos\theta \end{bmatrix} \tag{20}$$

Equation (20) has a pitch angle of θ, a roll angle of φ, and a heading angle of ψ, which can be realized as G_0 by equation (21):

$$G = \begin{bmatrix} 0 & 0 & -g \end{bmatrix}^T \tag{21}$$

$g = 9.81m \cdot s^{-2}$ in Equation (21), which can be calculated by Equation (22):

$$f^b = \tilde{f}^b - b_{a,e} - \eta_a \tag{22}$$

In Equation (22), the value obtained when performing accelerometer measurements is $\tilde{f}^b$ and the noise is η_a.

The gyroscope and accelerometer offsets are used for modeling to construct a first-order Markov model, as shown in Eqs. (23) and (24):

$$\dot{b}_g = -\frac{1}{\tau_g} b_{g,e} + \eta_g \tag{23}$$

$$\dot{b}_a = -\frac{1}{\tau_a} b_{a,e} + \eta_a \tag{24}$$

In Eqs. (23) and (24), the first-order Markov models of gyroscope and accelerometer bias are $\dot{b}_g$, $\dot{b}_a$, and the correlation time is denoted by τ_g, τ_a, respectively. Gaussian white noise is denoted by η_g, η_a in turn.

(2) Micro-inertial sensor fusion attitude fixing based on IU-EKF algorithm

Using the above pose estimation model and combining iterative update extended Kalman filter (IU-EKF) algorithm, the following steps are taken to realize the stance fixation of piano playing gesture:

(a) When the posture estimation measurement data z_k is acquired, this paper performs N steps of updating the measurement data in pseudo time, setting $N = 5$, at this time, the Kalman gain at each update is shown in Eq. (25) at $i = 1 \to N$ time:

$$K_k^{(i)} = \frac{1}{N}(p_k^{(i-1)} H_k^{(i)T} + c_k^{(i-1)})(w_k^{(i)})^{(-1)} \tag{25}$$

In Equation (25), each parameter is given the following definition: $w_k^{(i)}$ is the Jacobian matrix of the state vector, $H_k^{(i)}$ is the Jacobian matrix of the distance measurement function, $p_k^{(i-1)}$ is the Jacobian matrix of the input noise, and $c_k^{(i-1)}$ is the system noise covariance matrix:

$$w_k^{(i)} = H_k^{(i)} p_k^{(i-1)} H_k^{(i)T} + R_k + H_k^{(i)} c_k^{(i-1)} + c_k^{(i-1)} H_k^{(i)T} \tag{26}$$

$$H_k^{(i)} = \begin{bmatrix} \dfrac{\partial h_1}{\partial q_k^{(i)}} & 0_{3\times3} & 0_{3\times6} \\ \dfrac{\partial h_2}{\partial q_k^{(i)}} & \dfrac{\partial h_2}{\partial v_k^{(i)}} & 0_{3\times6} \end{bmatrix} \tag{27}$$

In Eqs. (26) and (27), R_k is the measurement noise covariance matrix, $v_k^{(i)}$ is the measurement noise, $q_k^{(i)}$ is the process noise, and h_1, h_2 are the transfer functions of $q_k^{(i)}$ and $v_k^{(i)}$.

(b) After updating the i th step measurement, the model state a posteriori estimate is shown in Eq. (28):

$$\hat{x}_k^{(i)} = x_k^{(i-1)} + K_k^{(i)} \left(y_k - h\left(\hat{x}_k^{(i-1)} \right) \right) \tag{28}$$

$h(\cdot)$ is the measure function of the nonlinear system, $\hat{x}_k^{(i)}$ is the state estimation vector, and y_k measures the noise variance.

The a posteriori error covariance is shown in Eq. (29):

$$\begin{aligned} p_k^{(i)} &= \left(I_{n\times n} - K_k^{(i)} H_k^{(i)}\right) p_k^{(i-1)} \left(I_{n\times n} - K_k^{(i)} H_k^{(i)}\right)^T \\ &+ K_k^{(i)} R_k K_k^{(i)T} - \left(I_{n\times n} - K_k^{(i)} H_k^{(i)}\right) c_k^{(i-1)} K_k^{(i)T} \\ &- K_k^{(i)} c_k^{(i-1)T} \left(I_{n\times n} - K_k^{(i)} H_k^{(i)}\right)^T \end{aligned} \tag{29}$$

where $I_{n\times n}$ is the system discrete state matrix.

(c) Perform steps a and b repeatedly until $i = N$, at this moment, the a posteriori state estimate is $\hat{x}_k^{+}$, and the a posteriori state estimate at the k moment is $\hat{x}_k^{(N)}$, meanwhile the a posteriori error covariance estimate is p_k^{+}, and k moment this value is $p_k^{(N)}$.

III. B. Piano playing gesture feature modeling and extraction methods

In order to extract the features of gesture data collected by inertial sensors, this paper investigates the multimodal feature extraction approach for playing gestures. An infrared detection rod is installed during the piano playing process, and the hand gesture data detected by the detection rod is extracted and used as auxiliary information for feature extraction. As a result, the features extracted in this paper are as follows:

(1) Statistical features related to the time domain: when playing the piano, the player's finger movements will change significantly, and the changes in the fingers will also change the movement amplitude of the back of the hand during the performance, in order to analyze the changes in the hand movements from multiple perspectives, this paper extracts the standard deviation of the hand gesture gesture, the extreme deviation, and the difference between before and after the pressing of the keys.

(2) Features based on spatial characteristics: the dynamic information of the fingers and the back of the hand during the player's performance is extracted to obtain the difference change of the posture angle between the back of the hand and each joint of the fingers when pressing the key.

(3) Coupling features between fingers: when performing daily performance, there are some differences in the changes between different fingers of the player, for this reason, this paper extracts the acceleration, angular velocity and other data between neighboring fingers.4) Auxiliary features: the use of infrared detector rods can effectively detect the keystrokes of each key in real time, and as a result, the finger movements during the current performance can be analyzed and derived on the basis of this information. Since it takes a certain time interval for the hand to make a movement when performing a performance, this paper realizes the management of the detection data by means of a sliding window with a fixed time width, and sets the width of the time window to 100ms, and extracts the data features by means of the width of the window, and uses the data as an auxiliary feature of the hand gesture. In this paper, the above features are normalized so that these features can be better applied in the recognition process, and are expressed by Equation (30):

$$p_{new-i} = \frac{p_{new-i} - p_{\min}}{p_{\max} - p_{\min}} \tag{30}$$

where p_{new-i} denotes the result of the normalization process, $p_{\max}$ denotes the maximum value of the feature, $p_{\min}$ denotes the minimum value of the feature, and p_{new-i} denotes the feature dimension. With this piano playing gesture feature modeling and extraction method, the recognition of piano playing gestures can be achieved.

IV. Piano technique analysis and optimization

IV. A. Effectiveness and analysis of finger rotation speed and angle prediction

IV. A. 1) Maximum angle results and analysis of the three paths of a finger touching a key

The experiment adopts the IU-EKF algorithm to realize the fixed posture of piano playing gesture, obtains the player's playing gesture posture, takes this playing gesture information as a data sample, and collects the piano playing gesture posture data by using MEMS inertial sensors. And the gesture posture estimation is made by the state space model. Based on this model, using the multi-feature extraction method, the gesture features are obtained, and different features are normalized, and the processed results are inputted into the Extreme Learning Machine (VGG-16) network model, and the recognition of the piano playing gesture is realized through the deep

migration learning and training of this model. This experiment focuses on accurately recognizing and analyzing piano gestures through MEMS inertial sensors. The focus was on verifying the accuracy of piano finger angles, for which the study designed three angle standard blocks to stabilize the bending angle of the fingers during playing. By analyzing the captured keypoints, the accuracy of each angle measurement of the finger is shown in Figure 3. The mean values of finger angles that can be calculated are 147.92°, 136.03° and 117.15°, respectively. The error range of these results is extremely small, with a maximum error of only 2.73% and a standard error of 2.62%, showing the high accuracy of the VGG-16 network model in recognizing finger movements.

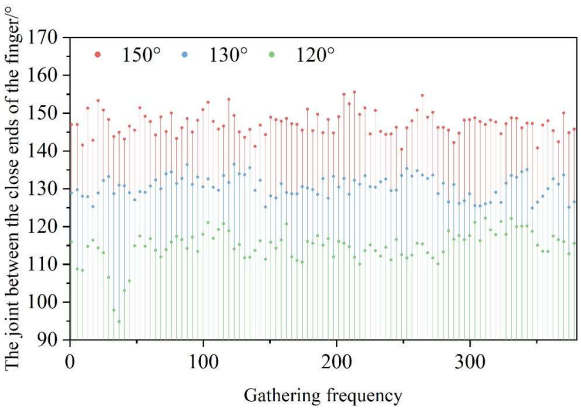


Figure 3: The accuracy of measurement in every Angle of the finger

When constructing the touch key model, the wrist is assumed to remain stationary. The base of the palm may produce a small rotation on the top glass. The central check is to determine whether the normal to the finger plane is parallel to the xy plane. This is investigated by obtaining the coordinates of the fingers, paying particular attention to the positions of the proximal interphalangeal joints, the distal interphalangeal joints of the fingers, and the fingertips. The angle between these coordinates and the line connecting the center of the palm to the xy plane was observed. The angular comparison of the three paths is shown in Figure 4. The maximum angular difference is only 2.6 degrees, which implies that minor wrist rotations can be omitted during keystrokes.

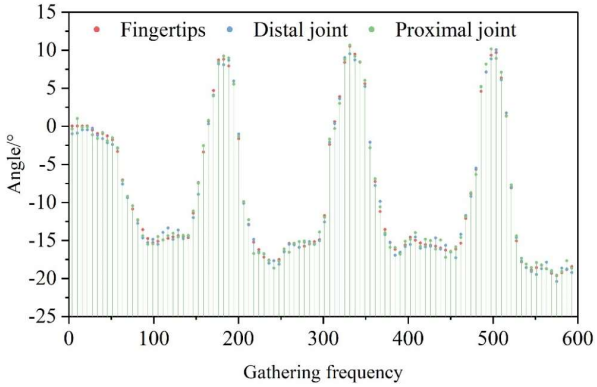


Figure 4: The Angle of three paths

IV. A. 2) Finger Speed Rotation Movement Results and Analysis

The movement characteristics of the finger in playing are shown in Fig. 5, (a) and (b) represent the rotation angle, and the curve of rotation angle with time after de-noising, respectively. A small change in the angle of the finger from the initial to 0.7 seconds indicates that the finger is at the bottom. Between 0.7 and 1.25 s, the data became sparse and the angle of the finger gradually increased, indicating finger lifting. 1.25 to 1.55s, the data stabilizes and the finger stops at the top end. 1.55 to 2.2s, fingers begin to slide as the angle of the palm joint decreases. At 2.2 s, stabilization was restored after a slight movement of the finger. In order to more accurately describe the keystroke velocity, the study removed the stabilizing portion at both ends and focused on the rising and falling velocity changes, see Fig. (b).

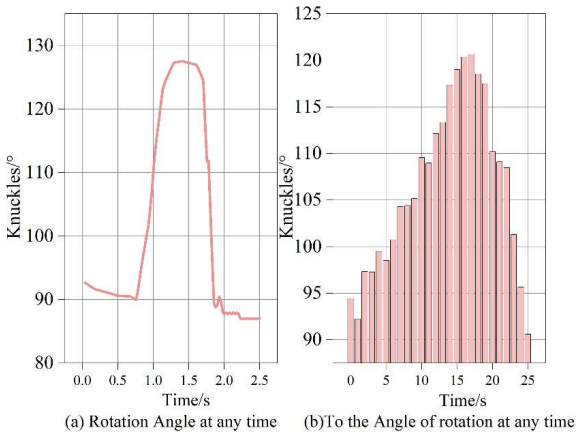
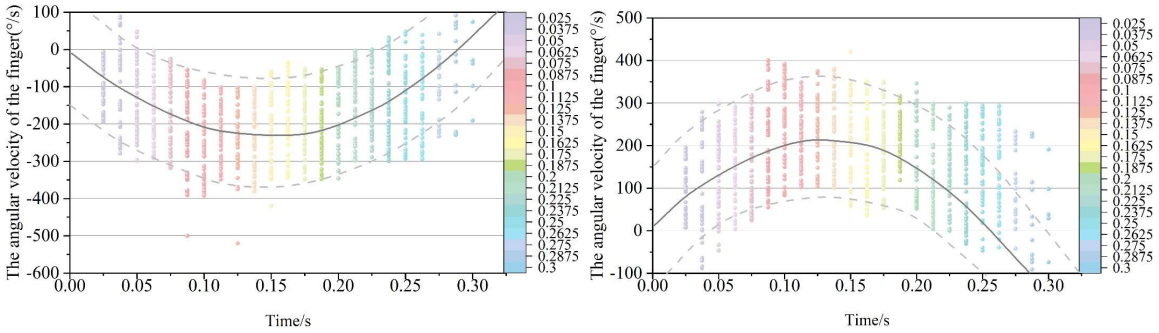


Figure 5: The motion characteristics of the fingers in playing

Derivation of the cycle movement versus time yields a velocity function, but the non-fixed structure of the hand makes each keystroke slightly different, so this analysis may be incomplete, while fitting the velocity to multi-cycle data is more complicated. To address these issues, the study introduces an innovative velocity calibration method. This method first estimates the average velocity using angular variations, then extracts the sliding curve from 35-60 movements and sets its onset time to 0 in order to focus on studying the single sliding velocity. After that, the palm joint rotational speed and duration matching between finger sliding and rebounding were studied. The change of the palm joint rotation speed of the finger during the slide and rebound phases is shown in Fig. 6, and (a) and (b) represent the palm joint rotations of the finger during the slide and the finger during the rebound after fusion, respectively. Fig. (a) demonstrates the matched rotational speeds under the 95% trust region, while Fig. (b) analyzes the finger rebound speeds in this same trust interval.



(a) The hand of the finger is rotated by the palm of the hand

(b) The fingers rebound when the palm of the hand rotates

Figure 6: The rotation velocity of the knuckles of the finger and the rebound stage

IV. A. 3) Predictive Effectiveness and Analysis of Musical Finger Skill Assessment Models

This research was trained on music sample data based on VGG-16. Each of these samples, has 14 input features with 1 output feature. During the training phase of the network, certain samples were used as the training set while others were the test set. The training set was primarily intended to construct an estimation model of musical piano technique, while the test set was used to test the performance of the constructed model. The size of the test set should be moderate when the overall sample size is limited. This inquiry used 90 samples as the training set and used 10 as the calibration set. After training the music samples based on these data, the actual and expected results of the test set are shown in Figure 7. The prediction effect of the trained neural network on the test samples is demonstrated, in which the test values are highly consistent with the actual values, and the network effectively estimates the finger assessment index.

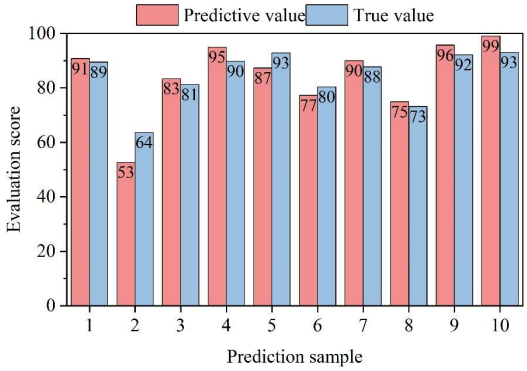


Figure 7: Comparison of the music sample test set

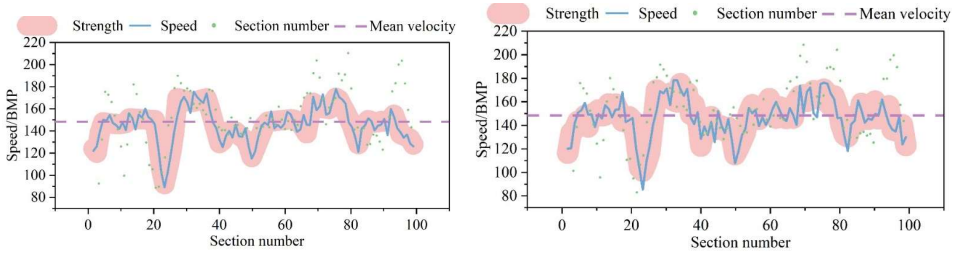
IV. B. Case Studies

Adapted from Goethe's narrative poem of the same name, The Demon King has clear and contrasting roles in the lyrics, and the composer creates a melody that matches the roles according to the lyrics, and the performer adds his own understanding of the music and the roles in the second composition by studying the score, which finally forms the work in the listener's ears.

In this chapter, different performance versions of The Lord of the Rings are visualized and analyzed based on computer vision technology to compare and analyze the similarities and differences in tempo-intensity processing between Petry and Bellman, with a view to extracting the differences in playing styles and the similarities and differences in tempo-intensity processing of the performers, and to provide theoretical basis for the optimization of piano technique training.

IV. B. 1) Overall speed-strength analysis

The overall velocity-intensity contrast is shown in Figure 8. (a) and (b) show the velocity-strength trends of Petry's and Bellman's performances, respectively. The nodes with large changes in velocity-strength are essentially the same, and the performers have similar velocity curves for the same passages. In terms of intensity, the change in intensity of the players is represented by the gray field on the velocity-intensity graphs, respectively. The larger the width of the gray area, the greater the intensity. As a whole, the intensity contrasts between the two players are handled in roughly the same way, with a similar distribution of strong and weak parts. The Petry version is smaller in volume than the Bellman version due to the limitations of recording technology, but the distinction between strong and weak parts can still be clearly seen. In the Bellman version, the distinction between the strongest and weakest parts is the greatest, and even in the strongest parts there is still a lot of subtlety in the strengths and weaknesses.



(a) Petri (b) Belman

Figure 8: Overall speed - strength contrast

IV. B. 2) Local speed-strength analysis

This section analyzes the tempo-intensity of the two performers' versions from a local microcosmic point of view. The piece was sectioned according to the schematic analysis, and the sectioning times of the performers are shown in Table 1.The proportion of time spent on each section is mostly similar between the two performers, but the time spent on each section is different. Bellman's version is not much slower than Petry's version at the beginning, and even faster than Petry's version in the introductory section and one paragraph. Starting from the second paragraph, when the narrator retires and the main characters of the story come out one after another, Bellman weakens the demonstration of technical skills and pays more attention to the portrayal of the characters.

Table 1: Graded schedule

Paragraph	Petri	Belman
Prelude	25.361	22.364
1 paragraph	37.643	35.946
2 paragraph	38.641	40.691
3 paragraph	27.631	38.9431
4 paragraph	22.569	31.424
5 paragraph	19.064	22.056
6 paragraph	27.619	29.061
7 paragraph	26.034	29.412
Epilogue	28.913	31.604
Recit	17.614	21.094

IV. C. Optimization Measures for Piano Technique Training

IV. C. 1) Enhancement of piano fingering practice

Fingering is the basis for practicing complex techniques, and piano practice should pay particular attention to fingering practice, not only to ensure that fingering is practiced according to a plan, but also to strengthen the reflection on fingering practice. First of all, the practitioner should try to figure out the fingering problems by himself, and develop a correct habit of fingering. For example, fingering practice should not be fast, it should be done to achieve the full rise and fall of the fingers, often finger relaxation exercises, control the power of each finger strike, pay attention to the basic fingering practice. Secondly, fingering practice can be used to video analysis and recording methods, video recording of their own piano playing, so that you can reflect more deeply, more targeted solution to the problem of piano fingering. Thirdly, piano practice can be used to break up the practice method, focusing on only one hand, you can effectively grasp the pitch, rhythm and fingering of one hand, analyze the problems in playing, so as to focus on breaking through the problems one by one. When the split practice reaches the point where you don't need to concentrate all your attention to play smoothly, then practice with both hands together, so you can achieve twice the result with half the effort. In the practice of some need to faster speed, greater strength of the repertoire, such as involving three degrees, octaves, ten degrees, decomposition of harmony and vibrato, etc., but also appropriate relaxation of the body and mind, to eliminate the tension of the hand muscles. If you practice for a long time and neglect to relax, it will affect the mental state of the player. Piano fingering practice should avoid the vicious circle, to prevent over-practice hand muscle damage.

IV. C. 2) Concentrate on technical difficulties

The main goal of piano technique practice is to overcome the technical difficulties and grasp the key techniques, so that after practicing for a period of time, the player is able to overcome certain aspects of the problem. First of all, the goal of the practice should be clearly defined, the method of practice should be defined according to the goal of the practice, and the practice process should be analyzed to see if the desired effect is achieved. Moreover, when breaking through the key techniques, they should not stop easily, and should adopt the practice method of catching up with the victory. Secondly, piano practitioners should also be good at deep understanding, actively analyze the gains and losses of practice, and deeply experience the emotion of the work when practicing, so as to reduce the waste of time and reduce the intensity of practice.

IV. C. 3) Actively drawing on the experience of others

Piano technique practice should also draw extensively on the experience of others, seriously summarize the gains and losses of other people's piano practice, learn from the master's piano practice methods, and improve the overall level of piano practice by constantly taking the best from the best. Piano practice is not only to figure out the sheet music, but also to spend a lot of effort to memorize the sheet music. Memorizing the music score can help you grasp the work as a whole and realize the refinement of the work. Before each piano practice, you should read the score carefully, try to understand the clef, key signatures, the position of each note, read the altered clef markings in the score carefully to prevent the problem of misremembering the key signatures, ascend or descend the notes at the right time, and memorize the position of the pitches in your mind, so that you can effectively avoid the phenomenon of playing wrongly or playing inaccurately. Piano practice is not only hands-on, but also with the heart to feel, through the auditory, visual cooperation, with the mind to practice, the realization of perceptual and rational cooperation, so as to improve the quality of practice, break through the disadvantages of a single practice, improve the effect of practice [17]. The so-called “brain practice” means practicing under the supervision of the brain and examining one's

own results from various aspects, so as to effectively avoid the problems of mechanical practice, fear of weaknesses in skills, and blind self-confidence.

V. Conclusion

The application of computer vision technology in piano teaching shows significant advantages and provides technical support for the change of traditional music education mode. The constructed multimodal gesture recognition system realizes the precise quantitative analysis of piano playing movements, the finger angle recognition accuracy reaches 97.27%, and the angular deviation of the three keystroke paths is controlled within 2.6 degrees, which proves the feasibility of the technical solution. The deep learning model performs well in gesture feature extraction, and the fusion of time-domain statistical features, spatial characteristic features, finger coupling features, and auxiliary features is used to reduce the standard error of playing action recognition to 2.62%. Comparative analysis of the performance versions of The Lord of the Rings reveals the individualized characteristics of different performers in the tempo-strength processing, and the Belman version reaches the maximum value in the contrast between strength and weakness, which provides a model for piano technique training. Based on the research results, the proposed optimization measures of strengthening fingering practice, focusing on technical difficulties, and learning from experience point out the direction for the reform of piano teaching in colleges and universities.

References

[1] Rottondi, C., Chafe, C., Allocchio, C., & Sarti, A. (2016). An overview on networked music performance technologies. IEEE Access, 4, 8823-8843.

[2] Webster, P. R. (2012). Key research in music technology and music teaching and learning. Journal of Music, Technology & Education, 4(2-3), 115-130.

[3] Malaschenko, V. O., Antonova, M. A., Knyazeva, G. L., Belokon, I. A., & Pechersky, B. A. (2020). Theoretical and practical aspects of integrating the computer and music technologies into the instrumental performance training of the pedagogical university students. In SHS Web of Conferences (Vol. 79, p. 01014). EDP Sciences.

[4] Lin, Y. J., Kao, H. K., Tseng, Y. C., Tsai, M., & Su, L. (2020, October). A human-computer duet system for music performance. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 772-780).

[5] Fabiani, M., Friberg, A., & Bresin, R. (2013). Systems for interactive control of computer generated music performance. Guide to computing for expressive music performance, 49-73.

[6] Yan, L. (2019). Design of piano teaching system based on internet of things technology. Journal of Intelligent & Fuzzy Systems, 37(5), 5905-5913.

[7] Ding, X., & Huang, N. (2022). Application of multimedia technology in online piano teaching. Mobile Information Systems, 2022(1), 1985546.

[8] Li, L. (2018). Application of augmented reality technology in piano teaching system design. Kuram ve Uygulamada Egitim Bilimleri, 18(5), 1712-1721.

[9] Niu, Y. (2021). Penetration of multimedia technology in piano teaching and performance based on complex network. Mathematical Problems in Engineering, 2021(1), 8872227.

[10] Yin, X. (2023). Educational innovation of piano teaching course in universities. Education and Information Technologies, 28(9), 11335-11350.

[11] Zheng, Y., & Wang, L. (2024). Application of entertainment virtual technology based on network information resources in piano teaching. Entertainment Computing, 50, 100675.

[12] Xue, X., & Jia, Z. (2022). The Piano‐Assisted Teaching System Based on an Artificial Intelligent Wireless Network. Wireless Communications and Mobile Computing, 2022(1), 5287172.

[13] Qian, L. (2024). Research and Practice on Instructional Methods for Piano Improvization Based on Computer Technology. International Journal of High Speed Electronics and Systems, 2440080.

[14] Wu Guobin & Chen Wei. (2022). Construction and Application of a Piano Playing Pitch Recognition Model Based on Neural Network. Computational Intelligence and Neuroscience,2022,8431982-8431982.

[15] Martin Kirchberger & Frank A. Russo. (2016). Dynamic Range Across Music Genres and the Perception of Dynamic Compression in Hearing-Impaired Listeners. Trends in Hearing,20,2331216516630549-2331216516630549.

[16] Sharma Sakshi & Singh Sukhwinder. (2021). Vision-based hand gesture recognition using deep learning for the interpretation of sign language. Expert Systems with Applications,182,115657-.

[17] Jing Guo,Jian Zhao,Xiaoxu Fan & Erpeng Song. (2019). The Value of Octave Technique in Promoting the Training Skills of Piano Performance. Поволжская Археология,30(4),