ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Assessment of the effect of piano playing gesture on sound quality based on motion capture technology

Shuyu Chen[1] *and Chi Zhang*[1,*]

[1] School of Music, Linyi University, Linyi, Shandong, 276000, China

SUMMARY: *In this paper, the micro-inertial sensor (MEMS) is used to capture the finger movement changes of the piano player, and the features extracted from them are used to form an improved multi-scale deep learning network to obtain better image fine-grained description ability and realize the recognition of the piano player's fingerwork. Finally, the perception and recognition function of piano playing gesture were realized by using the wearable piano playing glove. In addition, the speed of touching the key is analyzed to study its influence on sound quality. The experimental results show that our algorithm can solve the problems of low regularity, high variability and sudden change of hand movement pattern in piano playing, and has a recognition accuracy of higher than 99%. At the same time, speed and strength are the main control parameters, which directly affect the timbre change produced by the tapping technique, and all tapping techniques are in subtle changes to these factors.*

KEYWORDS: *MEMS; Motion Capture; Gesture Recognition; Piano Playing Gesture; Sound Quality*

1 Introduction

Piano performance teaching has a strong professional, need a lot of knowledge reserves to support, but in the traditional teaching, teachers one to one teaching method can not adapt to the economic burden of most students. At present, under the condition of insufficient number of teachers, face-to-face instruction for each student cannot be realized, so there is a large shortage [1-3]. Based on the students' performance data collected by wearable devices, autonomous learning was carried out to reduce the frequency of teachers' teaching. Teachers do not need to supervise students' practice at all times, but only need to view students' practice data on the Internet to understand all students' learning situation and practice status, and give more personalized and targeted teaching plans according to different students' practice data [5].

Computer aided instruction (CAI) refers to the use of computers to complete part of the teacher's work: to provide students with teaching materials, teaching methods and other information teaching means, is the product of the development of intelligent technology. With the development of information technology, it is more and more widely used in instrumental music teaching. For example, Linden, J et al. designed a violin bow instruction device MusicJacket, on which inertial sensors are installed to obtain the user's bow motion trajectory, compare this trajectory with the correct trajectory to obtain the error degree, and then guide students in bow method by means of haptic feedback [7]. K. Jakubowski and some other researchers used three different computation-based image processing methods: frame difference, optical flow, and kernel correlation filter (KCF) to build an effective tracking model

*zhangchibolin@163.com
https://doi.org/10.65102/is2026155 Published: 30 April, 2026

and demonstrated that the model can accurately capture the finger position of performers during music playing [8]. Secondly, Ashimori K, Igarashi H et al. designed a duplex tentacle glove device for performance practice, which can convert the teacher's hand movements into tactile feedback for students to help them master the learning process of finger playing skills on the guitar described in [9], imitating the teaching method. Is the way the teacher teaches with the body.

Many studies have shown that motion capture can record the body movements and hand shape changes of players in the playing process with the help of high-precision motion sensors and shooting equipment [10]. It can obtain more detailed information, such as the posture changes of fingertips, wrists, elbows, shoulders and other joints, or even the whole player [11, 12]. Through the analysis of the above information, the technical characteristics of piano players can be visualized, and the subsequent analysis work can be carried out on this basis. For example, Ameur, S et al. proposed a method of dynamic gesture recognition using Leap Motion equipment. They used LSTM neural network to perform machine learning on real-time time series data from Leap Motion, and applied it to other modules to form a new architecture. And provide a final gesture recognition prediction model [13]. In addition, effective methods can be used to detect students' gesture movements during piano performance, which is very useful guidance information for teachers [14].

For example, Xu and T combined CNN and RNN to propose a piano playing hand action recognition system, and proved that this method can effectively improve the accuracy of multi-action recognition and achieve real-time response effect. Of course, this also provides great help for intelligent piano teaching [15]. At the same time, Sha, R and Tan, B proposed the gesture recognition method of piano performance based on Leap motion capture and LSTM [16], which can well describe the movement of fingers touching the piano keys in the piano teaching process. Li, B et al. believe that expressive body posture is crucial for piano performance, and build a performance gesture recognizer based on deep neural network, which can recognize human gestures by training the movement patterns of specific players [17]. H Wang et al. combined CNN and GRU technology in the literature to design a set of finger recognition and finger training system based on motion capture technology, and it was mainly used for finger recognition of piano players. In this system, CNN was used to model the motion time series, and GRU was used to perform this work. They also added an attention module to the network to enhance the effectiveness and real-time performance of gesture recognition. R Wang et al. proposed a method that can effectively capture the action of playing the piano in the natural state by using unsupervised data sets and posture tracker, and combined with imitation learning and reinforcement learning methods to train a series of action rules for hand movements to be executed on the playing keyboard [19]. Sun and Y used the motion capture system to obtain the finger movement data of 8 piano professionals on the piano and constructed a function that conformed to the characteristics of finger touch key movement trajectory, which could provide accurate piano hand touch key trajectory for piano teaching [20].

From the above data, it can be seen that the application of motion tracking in music performance is gradually increasing, and it is also becoming more and more important for motion analysis of piano performance. However, compared with other gesture recognition applications, such as piano playing gesture recognition, there are great difficulties and challenges, including a wide variety of gesture categories, fast gesture transformation, large gesture change and obvious time dynamic characteristics. Therefore, the existing hand recognition methods are not applicable to this special scene, resulting in a series of problems [21-23]. For example, most hand recognition is based on static or stable gestures, which have obvious characteristics. However, for the piano playing gesture studied in this paper, it has the characteristics of fast speed, large amplitude and drastic change with time, which means that

we need to find a new model to cater to this special situation. And because of the large number of complex different hand shapes produced in the process of piano playing, and there is a very large similarity between them, the traditional method of finger recognition can not get good results.

The Extended Kalman filter based on iterative update (IU-EKF) is used to predict the piano player's hand position, and the hand motion information is used as the input. The piano player's hand motion information collected by MEMS inertial measurement unit is used to estimate the hand position by state space model. According to this process, the multi-channel feature extraction method was used to obtain the hand motion features, and the normalization processing was performed in each channel. An improved deep convolutional neural network classifier is designed to recognize hand gestures, realize the purpose of automatically extracting hand motion features from original images, overcome the shortcomings of traditional manual feature extraction, and screen out good scale features suitable for network model training. Finally, the effectiveness of the proposed method on piano playing gestures is evaluated, and the influence of piano playing gestures on sound quality is analyzed.

2 Piano playing gesture recognition algorithm based on motion capture technology

2.1 Piano Playing Gesture Recognition Techniques

The process of piano playing gesture recognition is as follows:

1) The small volume inertial measurement unit and infrared pen are used to collect the movements of the pianist, and the position information of each finger joint is obtained.

2) The signal emitted by the infrared pen is sampled by a window with a certain width and can be translated. Then it makes the score pictures of piano playing, and obtains valuable finger movement information.

3) Thirdly, the infrared scanning pen is used to collect the multi-modal feature information and complete the preliminary classification work. The multi-modal feature information is input into the time series model as the feature variable, and the correlation and spatial relationship are comprehensively considered, so as to establish the gesture representation model for recognizing the piano playing action.

2.1.1 Piano playing gesture estimation and stance fixation

(1) Piano gesture estimation based on state-space modeling

The hand data of piano playing gestures are collected by using MEMS inertial sensors. After the acquisition is implemented, the piano playing gestures are estimated by the pose estimation model.

The state space model is designed with the gesture of piano playing. In this paper, quaternions are chosen to describe gestures and the following parameters are used as sensor system state quantities:

$$x_e = \begin{bmatrix} q_e^T & v_e^T & b_{g,e}^T & b_{a,e}^T \end{bmatrix}^T \tag{1}$$

The unit quaternion of the gesture in the formula 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 $v_e = \begin{bmatrix} v_{east,e} & v_{narth,e} & v_{up,e} \end{bmatrix}^T$; the velocity components along the sky, east,

and north directions in the navigational coordinate system are denoted by $v_{up}, v_{east}, v_{north}$, respectively; the accelerometer offset is $b_{a,e} = \begin{bmatrix} b_{ax,e} & b_{ay,e} & b_{az,e} \end{bmatrix}^T$; gyro drift is $b_{a,e} = \begin{bmatrix} b_{ax,e} & b_{ay,e} & b_{az,e} \end{bmatrix}^T$. 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:

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

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, w_z, the elements of the unit-quadratic post-transpose matrix of the gesture pose are denoted by q_0, q_1, q_2, q_3, and the gyroscope output is denoted by $w = \tilde{w} - b_g - \eta_b, \tilde{w}$, and the gyroscope measurement noise is denoted by η_b. Jetlink inertial guidance specific force equation:

$$\dot{v} = R_b^n f^b + G_0 \tag{3}$$

The rotation matrix from the piano playing gesture coordinate system to the navigation coordinate system can be described by the unit quaternion R_b^n, the scale 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. Then there are:

$$R_b^n = 2\begin{bmatrix} 0.5 - q_2^2 - q_3^2 q_1 q_2 - q_0 q_3 q_1 q_3 + q_0 q_2 \\ q_1 q_2 + q_0 q_3 0.5 - q_1^2 - q_3^2 q_2 q_3 - q_0 q_1 \\ q_1 q_3 - q_0 q_2 q_2 q_3 + q_0 q_1 0.5 - q_1^2 - q_2^2 \end{bmatrix} \tag{4}$$

$$R_b^n = \begin{bmatrix} \cos\varphi\cos\psi + \sin\varphi\sin\psi\sin\theta \sin\psi\cos\theta \sin\varphi\cos\psi - \cos\psi\sin\psi\sin\theta \\ -\cos\varphi\sin\psi + \sin\varphi\cos\psi\sin\theta\cos\psi\cos\theta - \sin\psi\cos\psi - \cos\varphi\sin\psi\sin\theta \\ -\sin\varphi\cos\theta\sin\theta\cos\varphi\cos\theta \end{bmatrix} \tag{5}$$

The pitch angle is θ, the roll angle is φ, and the heading angle is ψ, which can be realized as G_0 by the following equation:

$$G = \begin{bmatrix} 0 & 0 & -g \end{bmatrix}^T \tag{6}$$

$g = 9.81\mathrm{m}\cdot\mathrm{s}^{-2}$, which can be calculated by the following equation:

$$f^b = f^b - b_{a,e} - \eta_a \tag{7}$$

The value obtained when performing accelerometer measurements is f^b and the noise is η_a.

Modeling using gyroscope and accelerometer offsets to construct a first order Markov model is obtained:

$$\dot{b}_g = -\frac{1}{\tau_g} b_{g,e} + \eta_{b_g} \tag{8}$$

$$\dot{b}_a = -\frac{1}{\tau_a} b_{a,e} + \eta_{b_a} \tag{9}$$

The first-order Markov model for the gyroscope and accelerometer bias is $\dot{b}_g, \dot{b}_a$, respectively, and the correlation time is denoted by τ_g, τ_a; the Gaussian white noise is denoted by η_{b_g}, η_{b_a} in turn.

(2) micro-inertial sensor fusion attitude localization based on IU-EKF algorithm

Combined with the above pose prediction model and the constantly updated Extended Kalman Filter (iu-ekf) algorithm, the state prediction of piano playing hand movements can be realized. The process is as follows:

① When the posture estimation measurement data z_k is acquired, in this paper, the measurement data is updated in N steps in pseudo time, and $N = 5$ is set, at this time, under $i = 1 \to N$ time, the Kalman gain at each update is as Eq:

$$K_k^{(i)} = \frac{1}{N}\left(p_k^{(i-1)} H_k^{(i)T} + c_k^{(i-1)}\right)\left(w_k^{(i)}\right)^{(-1)} \tag{10}$$

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:

$$w_k^{(i)} = H_k^{(i)} p_k^{(i-1)} H_k^{(i)T} + R_k + H_k^{(i)} c_k^{(i-1)} + c_k^{(i-1)} H_k^{(i)T} \tag{11}$$

$$H_k^{(i)} = \begin{bmatrix} \dfrac{\partial h_1}{\partial q_k^{(i)}} & 0_{3\times3} & 0_{3\times6} \\ \dfrac{\partial h_2}{\partial q_k^{(i)}} & \dfrac{\partial h_2}{\partial v_k^{(i)}} & 0_{3\times6} \end{bmatrix} \tag{12}$$

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 $v_k^{(i)}$.

② After updating the i-step measurements, the model state a posteriori estimates and error covariances are shown below:

$$\hat{x}_k^{(i)} = \hat{x}_k^{(i-1)} + K_k^{(i)}\left(y_k - h\left(\hat{x}_k^{(i-1)}\right)\right) \tag{13}$$

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

$$\begin{aligned} p_k^{(i)} = &\left(I_{n\times n} - K_k^{(i)} H_k^{(i)}\right) p_k^{(i-1)} \left(I_{n\times n} - K_k^{(i)} H_k^{(i)}\right)^T + \\ &K_k^{(i)} R_k K_k^{(i)T} - \left(I_{n\times n} - K_k^{(i)} H_k^{(i)}\right) c_k^{(i-1)} K_k^{(i)T} - \\ &K_k^{(i)} c_k^{(i-1)T} \left(I_{n\times n} - K_k^{(i)} H_k^{(i)}\right)^T \end{aligned} \tag{14}$$

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

Steps ① and ② are performed repeatedly until $i = N$. At this moment, the a posteriori state estimate is $\hat{x}_k^{(N)}$ at the moment of $\hat{x}_k^+, k$, and at the same time, the a posteriori error covariance estimate is $p_k^{(N)}$ at the moment of p_k^+, k this value.

2.1.2 Piano playing gesture feature modeling and extraction methods

In this paper, multi-modal gesture features are extracted, and various coupling features are extracted from gesture data in the process described in the previous section. And the infrared detection rod is used to obtain the gesture data measured by the detection rod in the process of playing the piano, as the information supplement of feature extraction.

In this paper, the extracted features are processed consistently in order to make better use of the features in the recognition process.

$$p_{new-i} = \frac{p_{new-i} - p_{\min}}{p_{\max} - p_{\min}} \tag{15}$$

p_{new-i} denotes the result of 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 dimension of the feature. Through this piano playing gesture feature modeling and extraction method, the recognition of piano playing gesture can be realized.

2.2 Improved Gesture Recognition Algorithm

2.2.1 Multiscale Convolutional Neural Networks

In deep learning networks, different sizes of convolutional layers can extract feature maps of different sizes. In order to detect target objects of different size intervals, we use adaptive multi-scale features, which mainly rely on convolution kernels of different shapes of the same type of convolution layer to generate features of different scales, and then concatenate low-order and high-order features to achieve the purpose of combining feature maps of different abstraction levels.

The multi-scale gesture recognition network structure proposed in this paper transmits the feature maps obtained from the ReLU activation layer in two ways: one is output along the normal conduction direction; The other one is directly output, and it is averaged and pooled before being sent to the fully connected layer. Finally, it is combined with the feature vector generated by this layer as the input of the Softmax layer for classification and recognition. The multi-scale CNN can extract different features for different levels of the image to achieve more accurate gesture classification and recognition, and reduce the amount of calculation used in

network optimization.

2.2.2 Selection of scale features

Because the change of gesture scale will seriously affect the recognition accuracy of the model, selecting the appropriate size feature is a key link in the construction of the convolutional neural network. In this paper, a greedy algorithm is used to select the size features. In order to avoid overfitting, the normalization constraint is added to the deep network on different size layers, and the best solution of each scale is obtained by using the supplementary introduced positive factor. The new regularity factors are:

$$C = C_0 + \frac{\lambda}{2}\sum_{\omega}\omega^2 \tag{16}$$

where C represents the new cost function, C_0 denotes the original cost function, λ is the regular parameter and ω is the corresponding weight. The following bias function can be obtained by taking the bias of the weights ω of the new cost function:

$$\frac{\partial C}{\partial \omega} = \frac{\partial C_0}{\partial \omega} + \lambda\omega \tag{17}$$

Therefore, the learning factor of the weights can be solved by transforming Eq. (17), i.e:

$$\omega' = \omega - \eta\frac{\partial C_0}{\partial \omega} - \eta\lambda\omega = (1-\eta\lambda)\omega - \eta\frac{\partial C_0}{\partial \omega} \tag{18}$$

It can be seen that the new weight update rule can be set to $1-\eta\lambda$, where η is the learning rate and $\eta\lambda$ is called the weight decay rate. By adjusting the size of λ, the corresponding size of the weights of the whole network can be changed.

2.2.3 Network training

To train the multi-scale convolutional network model proposed in this paper, all the forward propagation convolution formulas used in the paper are:

$$x_j^l = f\left(\sum_{i\in M_j} x_i^{l-1}k_{i,j}^l + b_j^l\right) \tag{19}$$

where l denotes the lth layer in the convolutional network; j denotes the jth kernel of the convolutional layer; M denotes the region where the convolutional kernel is located; k and b denote the convolutional kernel and bias, respectively; x denotes the value of the corresponding position of the feature map; and f denotes the activation function.

If we want to improve and update all the parameters of the fully connected network, we only need to calculate the output layer error and the hidden layer error during the backward conduction operation of the fully connected layer, and the specific expression is:

$$\delta_j = \left(d_{q,h} - x_{out,j}\right)g\left(x_j\right) \tag{20}$$

$$\delta_j^l = \left(\sum_{h=1}^{n^{i+1}} \delta_h^{l+1} \omega_{h,j}^{l+1} \right) g\left(x_j^l \right) \tag{21}$$

Eq. (20) is the output layer residual, and Eq. (21) is the hidden layer residual. $d_{q,h}$ denotes the corresponding desired output, $x_{\text{out},j}$ denotes the actual output, $g\left(x_j\right)$ denotes the derivative of the activation function, x_j denotes the output of the previous layer, and h, j denote the h th neuron and j th input. Therefore, according to the back propagation algorithm, the weights and bias of the fully connected network layers are updated as follows:

$$\Delta W^l = -\eta x^{l-1} \left(\delta^l \right)^{\mathrm{T}} \tag{22}$$

$$\Delta b^l = \delta^l \tag{23}$$

where ΔW^l denotes the weights of the l th layer, η denotes the learning rate, δ^j denotes the residuals of the l th layer, x^{l-1} denotes the output of the $l-1$ th layer, and Δb^l denotes the bias of the l th layer.

3 Assessment of the effect of piano playing posture on sound quality

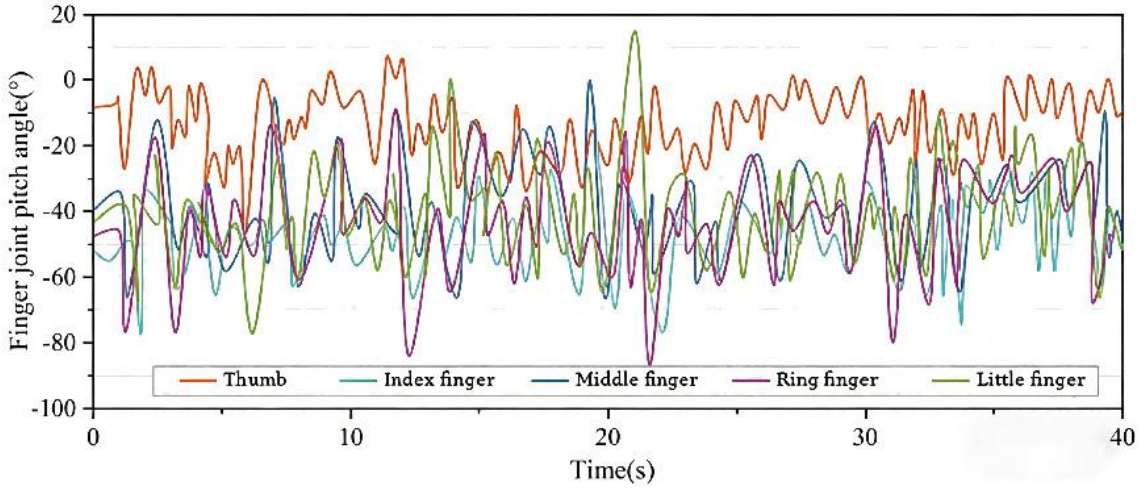
3.1 Analysis of Piano Playing Gesture Perception and Recognition Results

3.1.1 Piano playing joint integrity pitch angle change

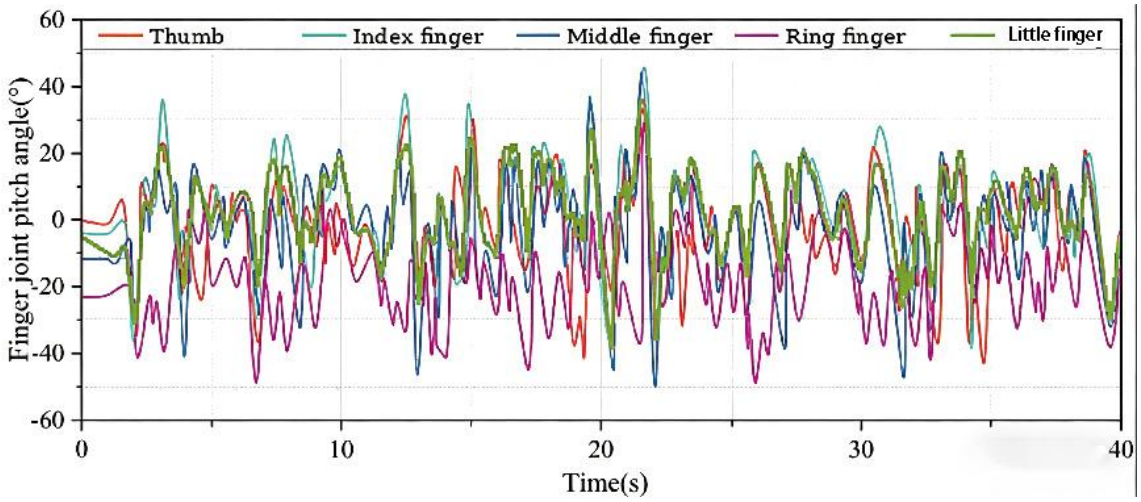
In order to verify the effectiveness of the proposed method, a wearable piano playing hand motion capture glove is used to perform intelligent perception of piano playing gestures, and the collection test work is carried out.

In this section, piano performances performed by 10 players were collected from "The Beautiful Dream God" and "Sonata in G major", and each piano work was sampled 4 times to obtain a total of 40 performance samples, which included data collected by the inertial data glove and data information collected by the infrared detection rod. In this part, the proposed algorithm will be used to test whether the designed inertial data glove can effectively collect hand piano performance movements.

Taking the sonata in G major as an example, the pitch Angle data of the upper joint of each finger is shown in Fig. 1, where (a) and (b) represent the upper and lower joints of the fingertip respectively. The proposed algorithm can better capture the changes of hand gestures, and the improved human gesture recognition algorithm is also better able to perceive the changes of hands.



(a)Metacarpal joint

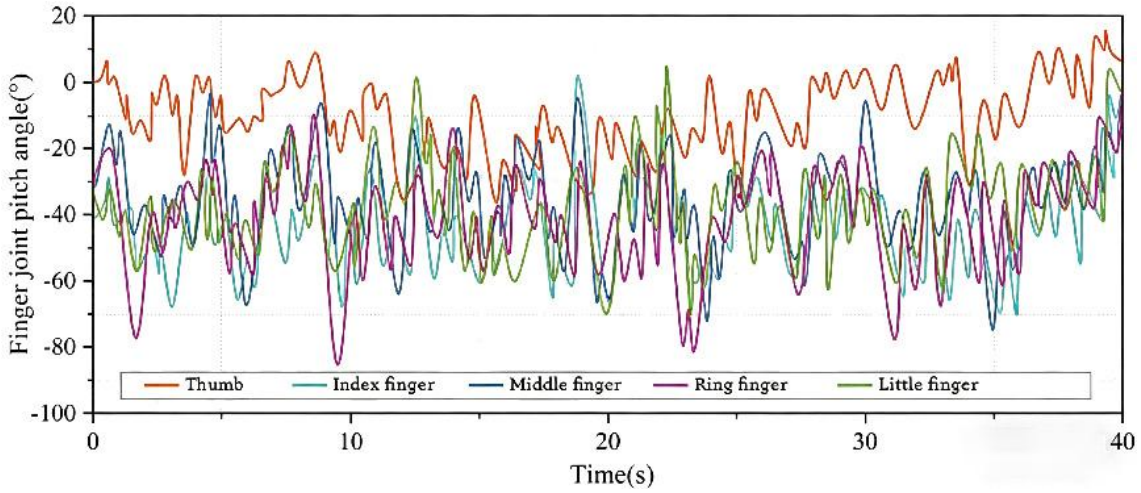


(b)Subdigital joint

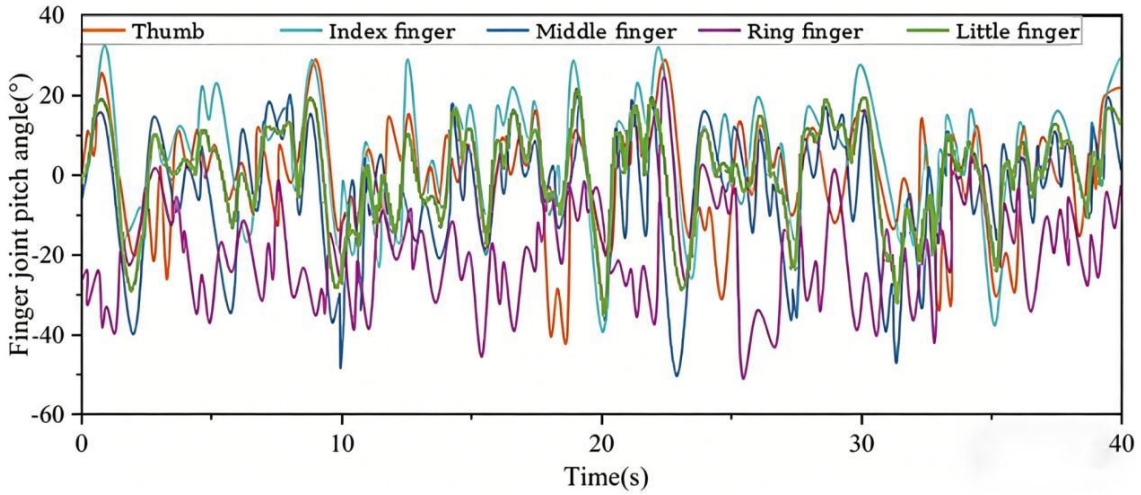
Figure 1: Pitch angles of the fingers' joints during playing

3.1.2 Pitch angle change of the piano joint at the moment of pressing a key

In this paper, the method of effective data segmentation and extraction is used to obtain the performance data of "Sonata in G major", and then the Angle change curve with time is drawn from these 40 effective hitting data, so as to obtain the movement category of the last performance. As shown in Fig. 2, after segmentation at the inflection point, it can be seen how the height Angle of the finger joints of the right hand playing the piano changes with time, specifically, including the joints above and below the fingers. In this way, the data acquisition can better capture the finger playing behavior. For the effect evaluation and analysis of our piano playing gesture recognition algorithm, it is found that this novel method can be used to build models and classify them.



(a) Metacarpal joint



(b) Subdigital joint

Figure 2: The result of the pitch angle variation of the piano playing joint

3.1.3 Piano playing hand movement recognition accuracy analysis

In this paper, four-fold cross validation is used to evaluate the performance of the designed recognition method. The 40 piano performance data collected are divided into four parts (each part has 10 different notes) according to the consistency, and four trials are carried out.

The method for hierarchical recognition algorithm proposed in this chapter and the traditional feature extraction method based on time series statistical information are compared and analyzed, and the recognition accuracy of finger movements of different piano players is shown in Table 1. It can be seen from the table that compared with the traditional method, the new algorithm proposed by us can get better recognition rate, and the recognition accuracy of finger pose judgment is increased by 25.09% compared with the traditional method, which greatly improves the accuracy of finger pose recognition.

Table 1: Recognition Accuracy of Piano Player's Hand Movement

Group	Conventional method (%)	The improved method in this paper (%)
Group 1	79.73	99.47
Group 2	80.67	99.66
Group 3	77.47	99.93
Group 4	80.85	99.62
Mean	79.68	99.67

The confusion matrix used in this paper is shown in Fig. 3, where 1 to 5 represent the predicted values of "thumb, index finger, middle finger, ring finger and little finger", respectively. A, B, C, D, and E denote the observations corresponding to each finger; 123, 135, and 1235 represent the gesture category with multiple fingers co-operating. The confusion matrix gives the expected classification by rows and the instances by columns. If the model has good prediction performance, all the examples can be accurately classified into the corresponding class, so they are in the main diagonal position, that is, the main diagonal is 1, and the rest are 0. Conversely, a poor model is associated with more misjudgments. Compared with 13 different confusion matrices, the improved recognition method can obtain better recognition and greatly reduce the confusion between hand movements.

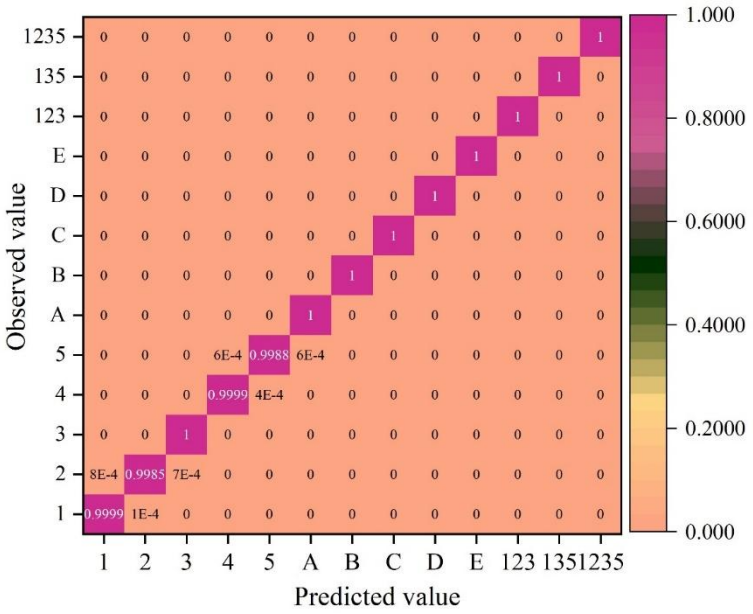


Figure 3: This paper improves the confusion matrix method

3.2 Influence of piano playing posture on the change of sound quality effect

3.2.1 Keystroke Velocimetry Experiment

Paste the touch metal plate on the button, and connect with the speed controller; The method in this paper is used to test the playing of the finger with a metal ring on the surface of the button touching the metal plate, and the two groups are compared to simulate the fast tapping and slow tapping. The actual situation of the data in the above two cases is displayed in the speed monitoring device; This process is repeated several times to ensure the validity of the data. At the same time, statistical methods are used to verify the above experimental conclusions. In this process, the data we get and the sound quality feedback from the audience in the actual

performance show that the speed of the tapping determines the size of the acceleration and the size of the vibration amplitude during the performance. In terms of controlling the strength and speed of the strike, we can appropriately increase or decrease the degree of the mallet acting on the strings, and then adjust the timbre.

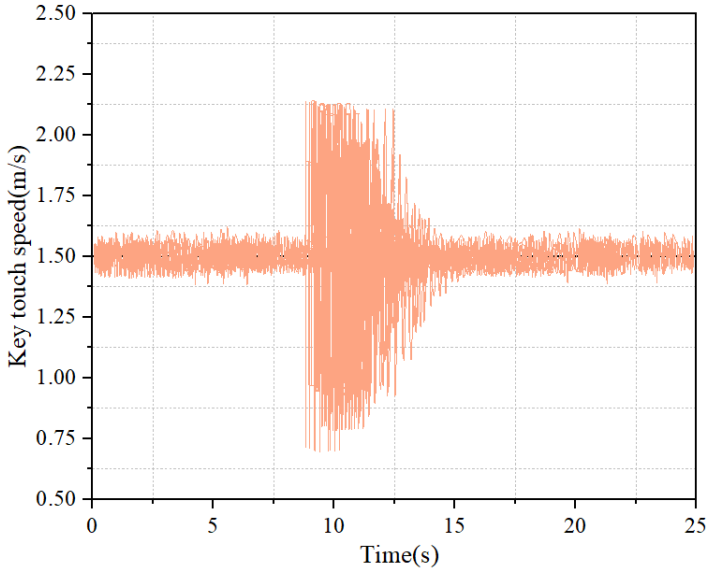
3.2.2 Experiments for analyzing the spectrum of touch keys

Through the spectrum analysis of two groups of keystroke experiments, this paper discusses the controllable factors that affect the change of the sound quality of keystroke in piano performance. Data analysis and comparison were carried out on the two kinds of key experiments of small role group A: "fast down key and immediately put key, slow down key and slow down key, and fast down key and then slow down key and immediately put key". In order to obtain real samples, two people play on a piano with the same tapping method in a fixed position to complete the experiment. This experiment was divided into two groups, named Experiment I and Experiment II. In this experiment, all experimental data and hand movement information were recorded as required by the experimental conditions. In order to ensure that the subjective feelings of the testers are fully expressed, and reflect the differences in individual cognitive level under different tapping techniques.

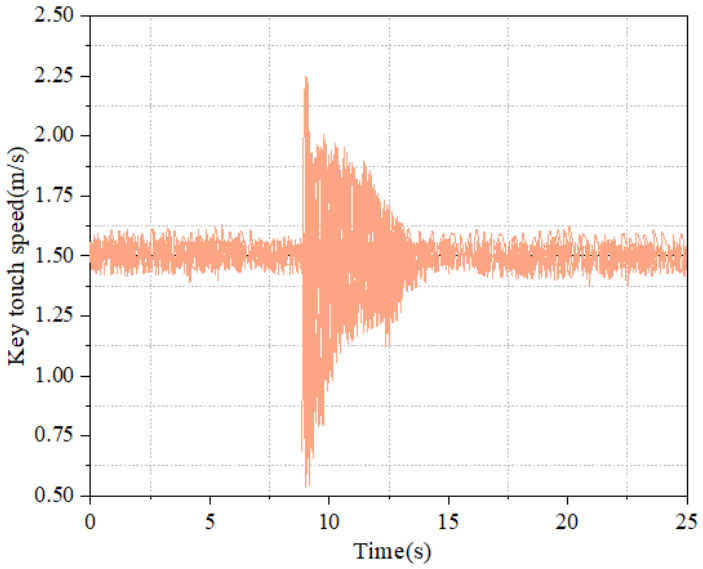
3.2.3 Experimental results

(1) Comparison of two groups of spectral data of fast bond dropping and fast debonding

Comparing the change of spectral value in the case of rapid key pressing and rapid key releasing, Fig. 4 shows the corresponding time-domain signal in experiment 1 in a. In b is the time domain signal corresponding to experiment 2; From the perspective of subjective perception, the perception of touching the button in experiment 1 was stronger than that in experiment 2. There was a continuous auditory sensation in Experiment two than in experiment one during the pull-down process. When we want to achieve a gravel-like effect, we tend to use the button method from Experiment 1. Through the observation of the spectrum, the audio with better image regularity and stronger sound tends to have more concentrated timbre. When the fundamental frequency is full and the spectrum shows a trend of regular decline in the audio frequency graph, the overall time-frequency graph is also more tidy. At this time, the sound cohesion and clarity of the syllable are stronger; On THE contrary, if the fundamental frequency is more powerful than the other parts and outside the amplitude of the spectrum, and the spectrum can be regularly attenuated in a short time, the speech is more discriminative (that is, different degree).



(a) Experiment 1: Time-domain waveform



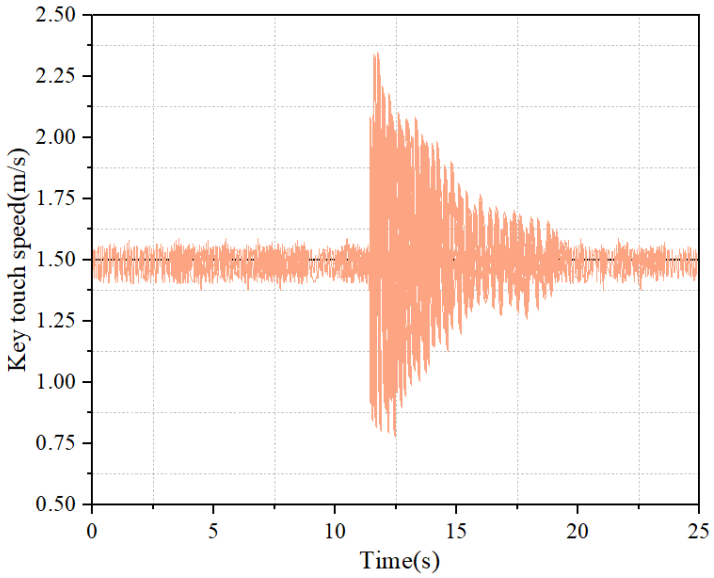
(b) Experiment 2: Time-domain waveform

Figure 4: Quickly press and release the keys to generate two sets of spectrum data

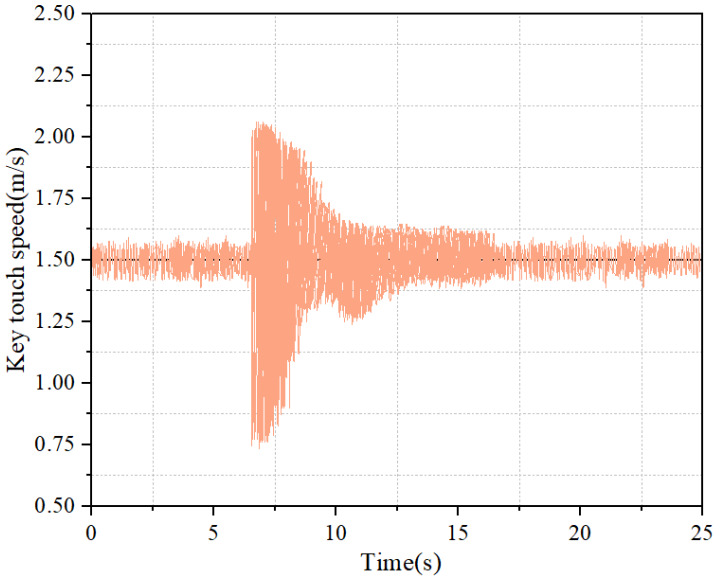
(2) Comparison of two groups of spectral data under fast bond and slow bond

See Fig. 5 for the comparison of the frequency values of the two groups of fast descending and slow descending presses the actual test waveform in Experiment 1 is (a); The actual test waveform of Experiment 2 was (b). In terms of practicality, the required sound quality should have concentrated good tone persistence, and Experiment 1 was better than Experiment 2 in terms of auditory perception and cognition. It has a much higher fundamental frequency. As for the problem of off-tone, the effect of Experiment II is relatively well maintained. To get the sound effect of long hair ends, you should use a combination of the one-finger touch method and the two-finger off-string method. If the fundamental frequency strength is large, a full and crisp sound will be issued; If the spectral audio effect is gentle and soft, it will be warmer and closer to the music. When the finger is off-key, the soft stopper stops the string from vibrating,

so the off-key speed determines the duration of the next note.



(a)Experiment 1: Time-domain waveform



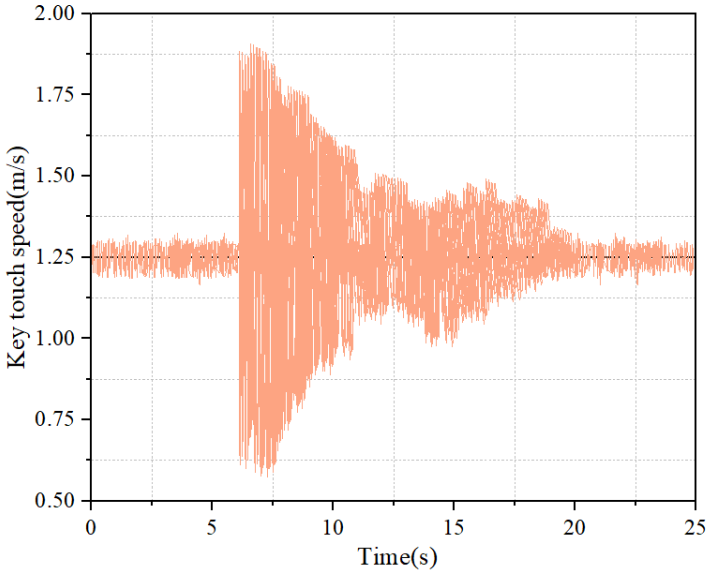
(b) Experiment 2: Time-domain waveform

Figure 5: Quickly press and slowly release the keys for two sets of spectrum data

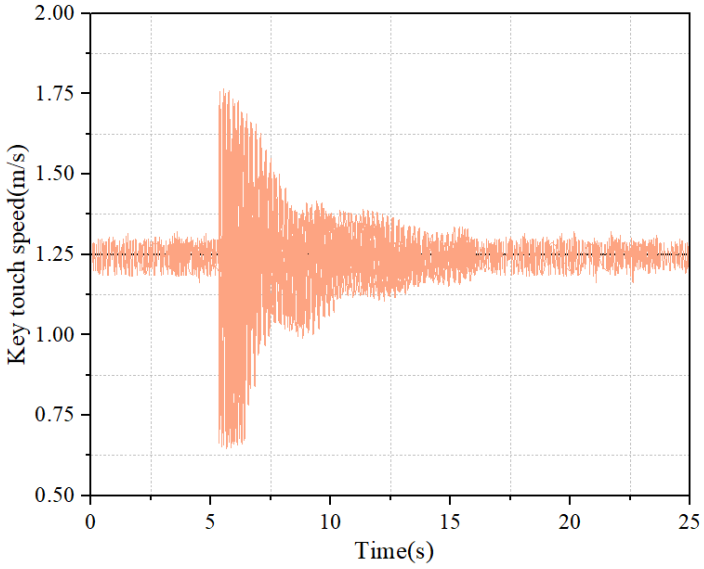
(3) Comparison of two groups of spectral data of slow descending bond and slow departing bond

Fig. 6 shows the frequency comparison plots of slow decline and slow separation for two different groups of controls, where (a) is Experiment I and (b) is Experiment II. Here, we expect to get a soft and continuous percussion timbre, that is, rich frequency components and good continuity, to avoid generating too strong fundamental frequency. This key pattern is often used in lyric songs, because in these songs, the individual characteristics of each tone are highlighted by weakening the overlap between them rather than allowing them to blend and lose their individual characteristics. In terms of perception, Experiment 1 was softer than Experiment 2,

which also performed slightly better for a period of time after the keystroke. However, if it is hard and fast to knock it down to generate a complete and clear fundamental frequency and spectrum, the generated sound will be more plump and bright. The fullness of the fundamental frequency depends on the speed of the intensity and depth when pressed. If you want to get a soft and continuous effect, the force of hitting the string should not be too large, so that the fundamental frequency is too sharp, nor too light. The phenomenon that the fundamental frequency amplitude is not enough and the sound is blurred.



(a)Experiment 1: Time-domain waveform



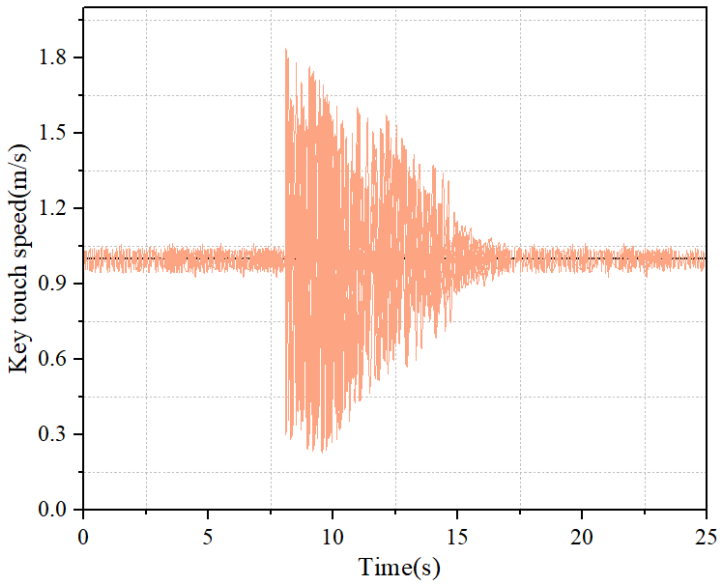
(b) Experiment 2: Time-domain waveform

Figure 6: Comparison of two sets of spectral data for slow-on and slow-off keys

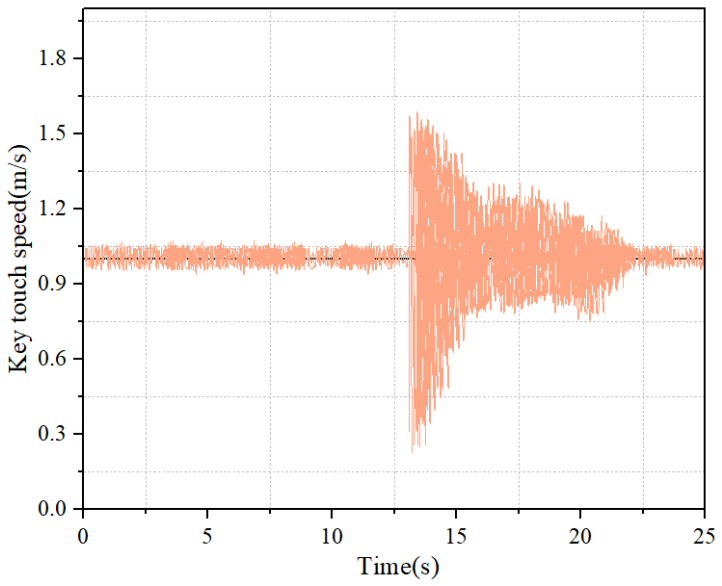
(4) Comparison of two groups of spectral data of slow bond decline and fast bond close

Fig. 7 shows the data comparison of slow frequency reduction and fast frequency reduction in two groups of different frequencies, that is, the time domain signals of Experiment 1 and

Experiment 2, as shown in (a) and (b), respectively. Because of the way of tapping and releasing the key, this method cannot be used for rapid continuous key pressing, so it is not a common typing mode. At the same time, this type of gesture is not often seen in classical music or other music, and is mainly used for onomorphic words or strong conflict. Such gestures require soft and short sound effects. In terms of tactile sensation, the stability of the first experiment was stronger than that of the second experiment, and the speed of the opening of the string was also faster than that of the second experiment. From the spectrum analysis, when the time series is more regular, the stability of the sound is higher, which indicates that the quality of the sound generated by experiment 1 is more stable than experiment 2.



(a)Experiment 1: Time-domain waveform



(b) Experiment 2: Time-domain waveform

Figure 7: Comparison of two sets of spectrum data with slow key and fast key

4 Conclusion

The piano playing gesture recognition method based on deep learning provides strong support for piano playing gesture recognition by collecting and extracting the features of piano playing gestures. Then, an improved multi-scale deep network gesture recognition model is proposed to analyze the influence of different types of piano playing gestures on the piano timbral. The specific results are as follows.

(1) The method proposed in this paper can not only better capture the action of the hand but also sense the dynamic information of the hand, which has a high recognition rate. Compared with the traditional recognition method, the accuracy of the gesture judgment of the action is improved by 25.09%, which can effectively reduce the confusion between different gestures.

(2) The strength and speed are the key elements that influence the sound quality of piano playing hand movement. The changes of sound quality produced by different key touching methods are as follows: fast down key, fast up key, sound is round and strong granularity, lack of prolonged sound; Fast key down, slow arm lift, concentrated point, strong sense of overtone column hierarchy, strong speech coherence.

References

[1] Carey, G., & Grant, C. (2015). Teacher and student perspectives on one-to-one pedagogy: Practices and possibilities. British Journal of Music Education, 32(1), 5-22.

[2] Dumlavwalla, D. (2017). Transitioning from traditional to online piano lessons: Perceptions of students, parents and teacher. MTNA e-Journal, 8(3), 2.

[3] Guobin, Z., Suttachitt, N., & Charoensloong, T. (2025). Designing Innovative Online Lessons to Foster Piano Playing Skills for Beginners with No Prior Experience. Journal of Posthumanism, 5(2), 247-266.

[4] Wee, C. C., & Mariappan, M. (2021). Hardware Design and Development of Contactless Sensor System for Piano Playing. In Control Engineering in Robotics and Industrial Automation: Malaysian Society for Automatic Control Engineers (MACE) Technical Series 2018 (pp. 199-208). Cham: Springer International Publishing.

[5] Dalla Bella, S., & Palmer, C. (2011). Rate effects on timing, key velocity, and finger kinematics in piano performance. PloS one, 6(6), e20518.

[6] Kaleli, Y. S. (2020). The Effect of Computer-Assisted Instruction on Piano Education: An Experimental Study with Pre-Service Music Teachers. International Journal of Technology in Education and Science, 4(3), 235-246.

[7] Van Der Linden, J., Schoonderwaldt, E., Bird, J., & Johnson, R. (2010). Musicjacket—combining motion capture and vibrotactile feedback to teach violin bowing. IEEE Transactions on Instrumentation and Measurement, 60(1), 104-113.

[8] Jakubowski, K., Eerola, T., Alborno, P., Volpe, G., Camurri, A., & Clayton, M. (2017). Extracting coarse body movements from video in music performance: A comparison of automated computer vision techniques with motion capture data. Frontiers in Digital Humanities, 4, 9.

[9] Ashimori, K., & Igarashi, H. (2018, March). Complemental Learning Assist for Musical Instruments by Haptic Presentation. In 2018 IEEE 15th International Workshop on Advanced Motion Control (AMC) (pp. 175-180). IEEE.

[10] Rahman, M. M., Hossain, A. A., Rana, M. M., & Mitobe, K. (2013, May). Hand motion capture system in piano playing. In 2013 International Conference on Informatics, Electronics and Vision (ICIEV) (pp. 1-5). IEEE.

[11] Zhou, L. (2025). Wearable Technology in Piano Training: Improving Posture and Motion Precision with Biofeedback Devices Like Upright Go. Applied Psychophysiology and Biofeedback, 1-10.

[12] Cheng, M. (2018). Introducing motion-capturing technology into the music practice room as a feedback tool for working towards the precision of rubato. Journal of Music, Technology & Education, 11(2), 149-170.

[13] Ameur, S., Khalifa, A. B., & Bouhlel, M. S. (2020). A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with leap motion. Entertainment Computing, 35, 100373.

[14] Yang, J., Zhou, Y., & Lu, Y. (2023). Multimedia Identification and Analysis Algorithm of Piano Performance Music Based on Deep Learning. Journal of electrical systems, 19(4).

[15] Xu, T. (2025, February). Research on Action Recognition and Feedback System in Piano Teaching Based on Deep Learning. In 2025 International Conference on Digital Analysis and Processing, Intelligent Computation (DAPIC) (pp. 658-664). IEEE.

[16] Sha, R., & Tan, B. (2024). Piano teaching action recognition based on LSTM. In Intelligent Computing Technology and Automation (pp. 669-675). IOS Press.

[17] Li, B., Maezawa, A., & Duan, Z. (2018, September). Skeleton Plays Piano: Online Generation of Pianist Body Movements from MIDI Performance. In ISMIR (pp. 218-224).

[18] Wang, H. (2025). Real Time Piano Finger Recognition Using Convolutional Neural Networks and Gated Recurrent Units. Informatica, 49(6).

[19] Wang, R., Xu, P., Shi, H., Schumann, E., & Liu, C. K. (2024, December). Fürelise: Capturing and physically synthesizing hand motion of piano performance. In SIGGRAPH Asia 2024 Conference Papers (pp. 1-11).

[20] Sun, Y. (2024, June). Piano Performance Techniques Based on Finger Motion Feature Capture Guide the Robot Analysis. In 2024 International Symposium on Intelligent Robotics and Systems (ISoIRS) (pp. 209-213). IEEE.

[21] Wong, G. K., Comeau, G., Russell, D., & Huta, V. (2022). Postural variability in piano performance. Music & Science, 5, 20592043221137887.

[22] Johnson, D., Damian, D., & Tzanetakis, G. (2020). Detecting hand posture in piano playing using depth data. Computer Music Journal, 43(1), 59-78.

[23] Chen, Y. C., Lin, C. Y., Chiou, Y. C., Chen, Y. C., Chen, M. S., & Lin, J. E. (2024, March). Enhancing Piano Practice Techniques: A Deep Learning System for Front Sitting Posture and Side Fingering Recognition. In Proceedings of the 2024 International Conference on Innovation in Artificial Intelligence (pp. 78-85).