



Value evaluation of human motion simulation based on speech recognition control

Jing Ma¹ · Jing Han²

Received: 18 August 2021 / Revised: 3 December 2021 / Accepted: 6 December 2021 / Published online: 4 January 2022

© The Author(s) under exclusive licence to The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2021

Abstract With the popularity of sports wearable smart devices, it is no longer difficult to obtain human movement data. A series of running fitness software came into being, leading the nation's running wave and greatly promoting the rapid development of the sports industry. However, a large amount of sports data has not been deeply mined, resulting in a huge waste of its value. In order to make the data collected by smart sports equipment better serve the sports enthusiasts, thereby more effectively improving the degree of informatization of the sports industry, this paper selects the design and implementation of the human motion recognition information processing system as the main research content. This article combs the previous research results of human motion recognition information processing systems related to sports wearable intelligence and proposes a three-layer human motion recognition information processing system architecture, including data collection layer, data calculation layer, and data application layer. In the data calculation layer, different from the traditional classification algorithm, this paper proposes a classifier based on the recurrent neural network algorithm. The mechanical motion capture method mainly uses mechanical devices to track and measure motion. A typical system consists of multiple joints and rigid links. Inertial

measurement units are bound to the joints to obtain angles and accelerations, and then analyze the human body motion based on these angles and accelerations. From the perspective of optical motion capture, the Kinect somatosensory camera is researched, and the method of human motion capture based on depth images, and the principle and method of human motion information are analyzed. At the same time, research on the application of Kinect's motion capture data. As a deep learning algorithm, convolutional action recognition model has the characteristics of being good at processing long and interrelated data and automatically learning features in the data. It solves the defect that the traditional recognition method needs to manually extract the motion features from the data, the whole system structure is streamlined, and the recognition efficiency is higher. The overall evaluation is as high as 99.4%. It avoids the manual extraction of time-domain and frequency-domain features of time series data, and at the same time avoids the loss of data information caused by dimensionality reduction.

Keywords Voice recognition control · Human motion simulation · Human–computer interaction · Human body model · Speech recognition control

✉ Jing Ma
yx900312@126.com

Jing Han
Hj057325@126.com

¹ Department of Software, Handan University, Handan 056005, Hebei, China

² Department of Foreign Languages, Handan University, Handan 056005, Hebei, China

1 Introduction

With the development of human social productivity and the continuous change of social production structure, informatization is increasingly becoming an important means of optimizing production structure and coordinating production methods, and it has also become an important way to promote supply-side structural reforms and

coordinate production relations¹. As a universal form of combining humans and computers, information processing systems have been widely used in all aspects of scientific research and social production, and life. They have become important for all walks of life to obtain information, store information, process information, communicate information, and use information tool.

Hsu Jason e found that only voice recognition controls the simulation of human movement, and the residual body movement around the graft is not sutured. Although it can significantly reduce the pressure on the subacromial contact surface, it cannot limit the upward displacement of the humeral head (Hsu Jason et al. 2019). Yang C studied the automatic motion control of the wheeled inverted pendulum (WIP) model, which has been applied to a wide range of modern two-wheel vehicle models. First, the underactuated W model is decomposed into a fully actuated second-order dynamics, including the plane motion of the vehicle bird motion and yaw angle motion, and a passive (non-action) forcing the system to swing motion. NN scheme has been used for motor control of neurotoxicity. The model reference method was used and the reference model was optimized by finite time linear quadratic adjustment technique. The motion proved valid for theoretical analysis and a simulation study was carried out to develop the method (Yang et al. 2017). Fields BKK reported for the first time that single-beam reconstruction of PCL under speech recognition can better restore the anatomical function of PCL (Fields et al. 2019). Harris g analysis showed that the increased levels of inflammatory factors on the first day after control were related to the stimulation of debridement under voice recognition. On the 7th day after control, the levels of PGE2 and IL-1 β in the two groups gradually decreased; after the control, the levels of PGE2 and IL-1 β in the observation group decreased more significantly than those in the control group (Harris et al. 2019). Xiang L believes that series elastic actuators (SEA) have a series of advantages over rigid actuators in terms of human–computer interaction, such as high force/torque fidelity, low impedance and impact resistance. Although various SEAS have been developed and implemented in initiatives involving physical interactions with humans, relatively few control schemes have been proposed to address the dynamic stability and uncertainty of robotic systems driven by SEA, as well as the open safety issues of resolving conflicts. The problem of movement between humans and robots has not yet been systematically addressed. He proposed a new continuous adaptive control method for SEA-driven robots for human–computer interaction. The proposed method provides a unified formula for the robot responsible mode (where the robot takes the lead to follow the desired trajectory) and the human responsible

mode (where the human takes the lead to guide the motion) (Xiang et al. 2017).

Homsy C reported 2 cases of wide swelling of the supraspinatus muscle notch in the exercise part of the human body. After matching the width of the moving part with the wide swelling, the movement part is wide without sutures, and good results have been achieved (Homsy et al. 2020). Such believes that only repairing the labrum without controlling the wide swelling around the glenoid of the exercise site can also achieve better results (Suh et al. 2019). Wang W proposed an electromagnetic energy collector based on adjustable magnetic spring to collect vibration energy from human movement. The collector was modeled by Ansoft Maxwell software, and the optimal magnetic stacking mode was selected according to the voltage generated by simulation. The dynamic model of the energy collector is derived and the corresponding theoretical and numerical analysis is carried out to evaluate the performance of the proposed system. Experimental results show that the harvester is expected to generate power in the wide band frequency range under the sweep excitation at different acceleration levels. In experiments considering human movement, the impact of shoes on the ground and the swing of legs were studied by connecting the device to the lower limbs. The test results show that proper structural parameters such as equivalent mass and moving length can improve the performance of harvester (Wang et al. 2017). Therefore, this paper proposes the voice recognition control of human motion simulation for the severe and severe human motion simulation caused by exercise. Short-term energy refers to the energy value of the speech signal in a short time, which is related to the strength of sound vibration. In layman's terms, the size of a person's voice is proportional to the energy consumed during this period of time. The louder the sound, the greater the energy consumed; the smaller the sound, the smaller the energy consumed; Generally speaking, when a person is angry, surprised, etc., the energy he emits is greater than when he is calm, that is to say, the voice is louder. Therefore, the short-term energy characteristic parameters can be used to distinguish between voiced and unvoiced sounds. At the same time, short-term energy can also be used as a basis for distinguishing between silent and voiced when the signal-to-noise ratio is relatively high.

Kinect has a total of three cameras, the middle of the front is the RGB color camera, and the left and right sides are the infrared transmitter and the infrared CMOS camera. An array microphone system is integrated at the bottom of the Kinect fuselage for voice recognition. In addition, Kinect is also equipped with focus tracking technology, the base motor will follow the movement of the focus object. Motion capture refers to recording the motion information of the human body in three-dimensional space through

sensing equipment, and transforming it into abstract motion data, and then analyzing the human body motion based on these data, which can then be converted into control signals for interactive intelligent control and In virtual simulation. In this paper, the Kinect sensor is used to obtain three-dimensional data of the human body, and then the data of human bone movement is obtained, and the human body movement is obtained by skeletal tracking and converted into control signals.

2 Human motion simulation algorithm

2.1 Voice control

In the process of pronunciation, because the vocal cords and lips will cause the voice signal to be affected by the glottal excitation, the high frequency end of the average power spectrum of the voice signal will fall at a speed of 20 dB/dec above a certain frequency. In order to solve this problem, Pre-emphasis was introduced. After the voice signal is pre-emphasized, the frequency spectrum of the signal becomes smooth, which can increase the amplitude of the high-frequency formant, so that the spectrum analysis or channel parameter analysis is easier. Realize the control of voice during human movement (Lee et al. 2019; Chi Andrew, et al. 2017).

After the framing of the speech signal, there will be discontinuities at the beginning and end of each frame, so the more the frames are divided, the greater the error with the original signal, and the Fourier transform function is called for the speech signal after the framing. The high-frequency part of the leakage phenomenon will occur at the time; therefore, it needs to be windowed to reduce the leakage, and at the same time, the signal after the frame becomes continuous, and each frame will show the characteristics of a periodic function (Zhao et al. 2019).

Since the development of human motion simulation field, a variety of physics-based simulation methods have been discovered and used. The simulation system mainly uses physics-based modeling methods, data-driven methods, controller methods and reinforcement learning methods. The dynamic control kinematics method has always been an enduring method to study the characteristics, when there is a large amount of data, this method may be effective. Given a data set of a motion clip, the controller can select the appropriate clip to play, in a given situation, and then synthesize the motion at runtime. Human body modeling is divided into kinematics-based models and dynamics-based models. The kinematics-based method describes the changes of various parts of the simulation model from a geometric perspective, so it does not involve the physical properties of the built model and the forces

experienced by each part. Therefore, using this method, it is difficult to express the constraints between the end motion trajectory of the simulation model and other joints of the object, and it cannot accurately reflect the dynamic characteristics of the simulation motion (Cuéllar et al. 2017).

2.2 Simulation algorithm

This part of the algorithm is not comprehensive, I have made up. In the process of using traditional algorithms to train the weights of the convolution kernel, the learning rate has remained unchanged and too large. Each training is to modify the weight of the convolution kernel directly to subtract the error of the convolution kernel weight. It is easy to cross the extreme point during training, so the traditional update algorithm may cause the update of the convolution kernel weight to be unstable as the number of iterations increases. For this reason, this paper proposes an improved algorithm, which makes the change of the weight of the convolution kernel every time it decreases nonlinearly or linearly with the increase of the number of iterations, so that the learning rate gradually decreases with the increase of the number of iterations.

2.3 Kinect sensor

Motion capture refers to the use of sensors to measure, track and record the motion information of the real human body and convert it into abstract motion data. In recent years, a lot of researches on motion capture have been conducted at home and abroad. The methods of motion capture are becoming more and more perfect, and the number of motion capture systems is increasing day by day. At the same time, traditional motion capture data is mainly used to drive virtual character animations and reproduce real human movements. Now, motion capture data is widely used in sports, entertainment, intelligent control and other fields, and its applications are becoming more and more diversified (Ranalletta et al. 2018; Latarjet 2020).

Kinect sensor is the XBOX360 somatosensory peripheral peripheral officially announced by Microsoft at the 2009 E3 exhibition. Its appearance has subverted the traditional game field and better embodies the human-computer interaction concept of somatosensory interaction. Compared with traditional cameras, the Kinect sensor adds real-time motion capture, voice input and recognition, image recognition, group interaction and other functions, so that players can interact with games more naturally and bring users a more realistic and natural gaming experience (Carbone et al. 2018).

$$BEC^{(k)} = \partial_1 l_{bce}^{(k)} + \partial_2 l_{iou}^{(k)} + \partial_3 l_{ssim}^{(k)} \tag{1}$$

$$Status = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})^2} = \frac{n \sum_{i=1}^n \sum_{i \neq j} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \tag{2}$$

$l_{bce}^{(k)}, l_{iou}^{(k)}, l_{ssim}^{(k)}$ are binary cross entropy (BCE) loss function, cross union ratio (IOU) loss function and structural similarity (SSIM) loss function. I is the super parameter of each loss function. BCE loss function is the most widely used loss function in binary classification and segmentation (Godenèche et al. 2020; Lafosse et al. 2017a).

$$l_{iou} = 1 - \frac{\sum_{a=1}^H \sum_{b=1}^W SEG(a, b)GT(a, b)}{\sum_{a=1}^H \sum_{b=1}^W [SEG(a, b) + GT(a, b) - SEG(a, b)GT(a, b)]} \tag{3}$$

$$l_{ssim} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4}$$

$$y = \beta X + (\ln - \alpha W)^{-1} \varepsilon \tag{5}$$

$$\rho_k = \frac{2k}{k+1} + \left[\frac{1}{2} + \frac{1}{2k} \right] \left[\frac{c_2 - c_1}{3} \right]^2 + \frac{2(c_2 - c_1)}{3} \tag{6}$$

$$MRI = \frac{2|SEG \cap GT|}{|SEG| + |GT|} \tag{7}$$

$$y = \alpha W y + \beta_1 X - W \beta_2 X + \varepsilon \tag{8}$$

$$RVD = 100\% \times \left(1 - \frac{|SEG| - |GT|}{|GT|} \right) \tag{9}$$

Sound is produced by the propagation of sound waves. It is an analog signal. It must be converted into a digital signal before it can be processed by a computer (Ali et al. 2020). Generally, the analog signal is converted into a digital signal by sampling, but in order to prevent the high-frequency part of the sound signal from being distorted, an anti-aliasing filter (low-pass filter) should be connected before sampling, so that the signal can be limited in the broadband within a certain range, the sampling rate is guaranteed to meet the sampling theorem (Ernstbrunner et al. 2018; Ekhtiari et al. 2018).

$$\ln MRI_{it} = a_0 + a_1 du * dt + \sum_{i=1}^N b_j Xu + \varepsilon_u \tag{10}$$

$$(\ln - \alpha W)y = (\ln - \alpha W)X\beta + \varepsilon \tag{11}$$

$$P(d_i, w_j) = P(d_i)P(w_j|d_i); \tag{12}$$

$$P(w_j|d_i) = \sum_{k=1}^K P(w_j|z_k)P(z_k|d_i)$$

$$U_1 = \begin{cases} s - p_1 - x_1 \\ x - p_2 - (1 - x_1) \end{cases} \tag{13}$$

$$\ln Med_{it} = a_0 + a_1 du * dt * \text{agg} + \sum_{i=1}^N b_j Xu + \varepsilon_u \tag{14}$$

$$SEN = \frac{|SEG \cap GT|}{GT} \tag{15}$$

$$\theta(p, q) = \arctan \left(\frac{L(p, q + 1) - L(p, q - 1)}{L(p + 1, q) - L(p - 1, q)} \right) \tag{16}$$

$$\psi = \sum_{x=1}^{\theta} Vx = \sum_{x=1}^{\vartheta} \left(\frac{Wx}{\sum_1^n W_{\mathfrak{S}}} Sx \right) \tag{17}$$

$$\mathfrak{R} = \frac{|SEG \cap GT|}{|SEG \cup GT|} \tag{18}$$

$$x_H = \frac{p_2 - p_1 + 1}{2} \tag{19}$$

This article chooses the design and implementation of the human motion recognition information processing system as the main research content (Makhni et al. 2018). After the data of each frame is collected, data processing is required until a certain amount of data is collected. Save the collected human skeleton data as BVH format for output. The BVH file contains two parts, skeleton information and motion data. The skeleton information describes the hierarchical structure of the skeleton with the human hip joint as the root node, including information such as rotation, joint offset, and number of channels. The motion data records the joint Euler angle data captured in real time according to the skeleton hierarchy, which is convenient for human body motion modeling and motion generation (Lafosse et al. 2017b; Hurley et al. 2019).

$$x_W = \frac{p_2 - p_1 + k}{2k} \tag{20}$$

$$Goutlliar = 1 - \frac{|SEG \cap GT|}{|SEG \cup GT|} \tag{21}$$

$$U_2 = \begin{cases} s - p_1 - kx_2 \\ x - p_2 - k(1 - x_2) \end{cases} \tag{22}$$

Kinect is able to have powerful in-depth image acquisition and motion data capture functions, mainly because it has built-in somatosensory detection device PrimeSensor developed by Israeli PrimeSense company and sensor chip PS1080, these two hardware devices rely on light coding (LightCoding) Technology to capture the depth information of the current scene. The so-called optical encoding technology refers to the use of continuous light (near infrared) to encode the measurement space, and then the encoded light is read by the sensor, and then decoded by the chip operation to generate an image with depth information (Cerciello et al. 2018). The key of optical coding

technology is laser speckle. When the laser hits rough objects or passes through ground glass, random reflection spots will be formed, which is called speckle. Speckle has a high degree of randomness, and its pattern will also change with the distance (Owling et al. 2016; Beibei, et al. 2019). The speckles in any two places in the space will be different patterns, which is equivalent to adding a mark to the entire scanning space, so any object when entering the space and moving in the space, the position of the object can be accurately recorded. The so-called encoding of the measured space refers to the use of laser to generate this speckle. Kinect uses infrared rays to emit class1 lasers that are invisible to the human eye. Through the grating in front of the lens, also known as the diffuser, the laser is evenly distributed and projected into the measurement space, and then each speckle in the space is recorded through the infrared camera. Finally, it is calculated by the wafer into an image with depth information (Zhen-Bin et al. 2018).

2.4 Human motion model

At present, the research of simulation technology is mainly embodied in the simulation of human–computer interaction, the construction of human–computer virtual environment and human body model, which is also the main direction of this thesis. The market trend has brought the design and research of ergonomics to a new height and put it on the agenda. Therefore, it is urgent to make further breakthroughs in its research to improve the market competitiveness of enterprises. The research of system simulation of multi-rigid body motion has been relatively mature. The main methods include: Lagrange dynamics, Newton–Euler equation, variational method, Kane method and so on. Among them, Lagrange dynamic equation is a kind of system equation about constraint force. There are two main types of this method: the first type of Lagrangian equation and the second type of Lagrangian equation. The first category applies to complete systems and non-complete systems, and the second category only applies to complete systems. The generalized coordinates of this method require auxiliary constraint equations to obtain effective solutions. The Lagrangian dynamics function is simple and easy to implement with the program, and the dynamics provides the completed force mode, inertia matrix and system structure, but this method amount of calculation is relatively large. The topological relationship of the human body model system is generally composed of a set of rigid body models related to a certain structure, such as: head, body, limbs, hands, feet, etc., and the connection between rigid bodies can be made by springs one by one. The movement of each rigid body needs to be converted to generalized coordinates, which requires the use of the Jacobian matrix of space transformation, and the

degree of freedom of the human body in the generalized space—generally ranging from 10 to 200, depending on the accuracy of the human body model.

Motion modeling is the core of human motion simulation. The so-called establishment of a human body motion model is to describe the connected limbs in the human skeleton model and the movement relationship between these limbs. The joints are regarded as points, and the bones between the joints are regarded as chains (Xia et al. 2017; Lei and Li 2017). The limbs are linked to represent the human skeleton as a tree-like hierarchical structure (Siqueira et al. 2017).

The vast majority of human movements are done by limbs, so the main research object of this article is the movements of human limbs. The movements of human limbs are very flexible and complex. It is difficult to describe all movements with a unified mathematical model. Therefore, in order to describe the movements of the limbs, it is necessary to record the relative posture of the limbs in each frame of motion data. Since the movement of the limbs is basically formed by the rotation of the joints, pure displacement movement is not common. Therefore, the posture changes during the movement of the limbs can be described by the angle-axis posture description method (Siqueira et al. 2017; Jurkovic et al. 2016).

Obviously, the first mandatory component to construct a human model is the physics engine itself. Simply put, the goal of a physics engine to simulate a virtual object is to propose a credible trajectory. These virtual objects include non-deformable and inelastic rigid bodies and deformable soft bodies. By subjecting these objects to virtual forces such as gravity and friction, the physics engine produces believable motion. Objects can be connected to each other through constraints, effectively restricting their relative motion. In addition, the physics engine can perform collision detection and take into account the properties of the simulated object (such as elasticity, friction coefficient). A physics engine simulates with a time step. In each simulation step, four operations are performed. First of all, the conflict processing stage determines whether the entities exist with each other, and if so, resolve the conflict. In other words, when the geometry of the object intersects, a collision is detected. When such an event occurs, the simulation solver calculates the force required to resolve the contact. Second, apply force and moment to the simulated object. These forces can be external (such as gravity) or internal (such as joint constraints and muscle forces generated by the controller). Third, the forward dynamics phase uses Newton–Euler dynamics laws to calculate the linear and angular velocities of each simulation object. The calculated speed depends not only on the previously applied force and moment, but also on the mass and inertia tensor of the object. In addition, the forward dynamics

stage ensures that the contact and constraints between the rigid bodies are guaranteed. Finally, the numerical integration operation updates the linear and angular velocity, position and direction of the object by integrating the acceleration calculated in the previous step.

3 Human motion simulation experiment controlled by speech recognition

3.1 Development platform

The acquisition of depth images is the core of Kinect. Kinect uses PrimeScene's LightCoding depth measurement technology. The so-called LightCoding uses a light source to encode the required measurement space, which is a kind of structured light technology. The light source of LightCoding is laser speckles, which are diffraction speckles randomly formed after the laser penetrates the ground glass. These speckles have a high degree of randomness and will change patterns with different distances, just like marking the space and objects in the space. The position can be judged by the speckle pattern above. The infrared projector projects infrared light, and the infrared camera acquires the speckle pattern marked in the space. At intervals, a reference plane is taken, and the speckle pattern on the reference plane is recorded. After multiple speckle pattern cross-correlation operations and interpolations three-dimensional shape of the whole scene can be obtained by calculation.

The development platform of this research is the Unity3D development engine and the Microsoft Visual Studio 2012 development tool. A complete system with virtual character action simulation functions and a series of human–computer interaction functions is developed on the Unity platform. The main development languages used in this system are C++ and C# script development languages.

3.2 Speech recognition engine

Speech recognition is mainly composed of endpoint detection, feature extraction, and pattern matching. Endpoint detection is to accurately detect the starting point of the voice signal and distinguish the voice signal from the noise signal. Feature extraction is to find the parameter pattern matching that can effectively represent the voice feature (that is, to calculate the similarity between a segment of voice and the feature template). After the process of processing the original speech signal, the final recognition result is obtained.

3.3 Motion capture system design

The function of the remote control module is realized by the control button and the reset button. The function of the control button is to send control signals to the signal acquisition module and the PC receiving module. When the two modules receive the signal sent by the module, they will send their respective confirmation signals to the remote control module. After pressing the control button, if the reset button is not pressed, pressing the control button repeatedly is invalid. The function of the reset button is to make the control button effective. After the data is collected, the reset button needs to be pressed to start the second data collection. The acquisition and storage module is worn on the waist of the tested athlete and is used for real-time acquisition and storage of sports information. The MTI sensor is mainly placed in this module. There is a storage unit inside the module, and the movement information will be stored in this unit. After the data collection is completed, the data needs to be imported to the PC through the PC receiving module, and the internal storage unit of the module is cleared. The synchronization signal acquisition module is installed next to the camera, and the indicator light of the module should be placed in front of the lens to achieve synchronization with the camera, so that it is convenient to compare the data analyzed by the capture system and the video analysis data. The motion capture control process is shown in Fig. 1.

3.4 Voice control simulation exercise

Write a script file, call the implemented Baidu speech recognition method, obtain the text result of the speech recognition feedback, analyze the corresponding instruction, change the condition variable of the animation BTree by judging the password, realize the transition and transformation of the animation, and achieve the voice input control character model simulates the target of the movement.

After the animation and voice integration module is completed, the module functions are integrated on the unity development platform, and the voice control three-dimensional model is converted into simulation animation. Import the 3D model of the bound human skeleton in the Unity project, use the skeletal animation system, use the skeleton motion data frame to drive the 3D model, add related components, and complete the creation of the character model animation. Create a simple scene in Unity3D, add the code of the speech recognition technology module, get the feedback result after calling the speech recognition technology, compare with the animation control password, realize the 3D model animation conversion,

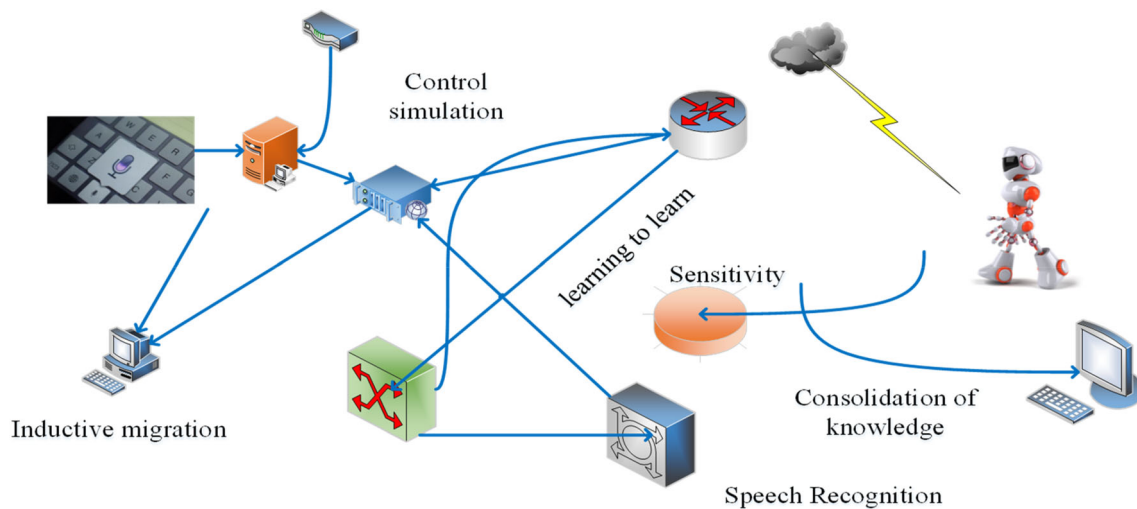


Fig. 1 The motion capture control process

in order to achieve the effect of voice control animation conversion.

4 Human movement recognition information processing system

4.1 Voice control motion simulation data analysis

Human body motion simulation is a technology that uses computers to simulate the motion process of real human bodies in nature, including the establishment of a calculation model, simulation and calculation of the physical motion process of a virtual human under a given constraint, and the corresponding three-dimensional graphics are generated on the computer to present the motion process. This technology can not only be used to enhance the authenticity and sensory nature of the virtual environment, but also can be used to simulate the escape behavior of the human body under natural disasters. Since the human body is a very complex structure, especially the bone structure, facial expressions, skin changes, etc., it is very demanding to simulate the changes in a human body in real time with the current hardware and software conditions, so it is necessary to model the human body model. Simplify the processing and use the corresponding physical formulas to constrain the changing process of the human body. Moreover, the interaction process with the outside world is too complicated to satisfy the real-time nature of simulation. It is necessary to simplify it and use certain technical means to deal with this interaction process.

The number of participants in this user evaluation is 121, mainly concentrated in the in-service personnel, including 41 in the professional field and 80 in the general

public. The total score is 100. The overall evaluation is as high as 99.4%. The evaluation results are shown in Table 1.

Edit the animation state machine, add variable conditions, and test the conversion feedback time of different animations by entering the voice passwords “Run”, “Run”, “Jump”, “Walk”, “Walk”, “Beckon”, and “OK”, and observe the differences for the fluency of transition between animations, the test results are shown in Table 2.

The recognition effect of the force curve corresponding to the movement phase of the same athlete is still relatively ideal. The six-dimensional foot force information recognition of a single athlete is shown in Fig. 2.

Figure 3 shows the identification of six-dimensional foot strength information of multiple athletes.

Human body motion simulation requires firstly establishing a human body motion model, and then building a dynamic system and simulation calculation on the motion model. Human motion simulation methods are divided into two categories: forward dynamics-based methods and inverse dynamics-based methods. Based on the forward dynamics method, the human body’s motion trajectory is calculated by specifying the driving force on the human joints and the external force acting on the human body by the environment, and integrating the human body mechanics system forward. Based on the inverse dynamics method, the human body motion initialized by the input is

Table 1 The evaluation results

Dimension	Voice driven	No voice drive
Interoperability	95.5	63.5
Realism	90.4	99.5
Overall evaluation	99.4	50.6

Table 2 The test results

Character animation type	System response time (ms)
Run	600
Walk	500
Jump	400
Wave	650
OK	450

optimized, so that the motion of the object meets the physical constraints and user-specified constraints. The steps of this method are to first input the initial human motion, and then enter a time sequence and set the motion restrictions. Finally, according to the set human motion dynamics equation, let the motion of the object meet the constraint condition. Figure 4 shows the test results after REST API and Android SDK are integrated into Unity.

4.2 Discussion

With the development of science and technology, human-computer interaction technology has been transformed from GUI to NUI. The so-called NUI is natural human-computer interaction technology. It uses the most natural methods such as voice, somatosensory, gestures, brain waves, etc. to realize computer interaction through

different sensors, etc. The control of intelligent machines has got rid of traditional interactive methods such as mouse and keyboard. Human motion simulation is a research field formed by the intersection of multiple disciplines and multiple technologies. It involves such disciplines as robotics, kinematics and dynamics, as well as technologies such as motion data capture, virtual human motion generation and control.

Human body recognition is based on depth images. Kinect performs pixel-level evaluation of depth images, first see the outline and then the details. Analyze the area closer to Kinect, which is also the most likely target of the human body. Any “large” font object will be considered by Kinect as a player; then scan the pixels of the depth image of these areas point by point to determine which parts of the human body belong to, including the computer graphic vision technology, such as edge detection, noise threshold processing and classification of human target feature points, distinguishes the human body from the environment. Kinect can actively track the whole body skeleton of up to two human bodies, or passively track the shape and position of up to six human bodies.

Human motion simulation is a field where multiple disciplines such as biomechanics, robotics, and computer science intersect. The main research content of robotics is robot kinematics and robot dynamics. The main function of the kinematics and dynamics simulation experiment module is research. And verify the kinematics and dynamics

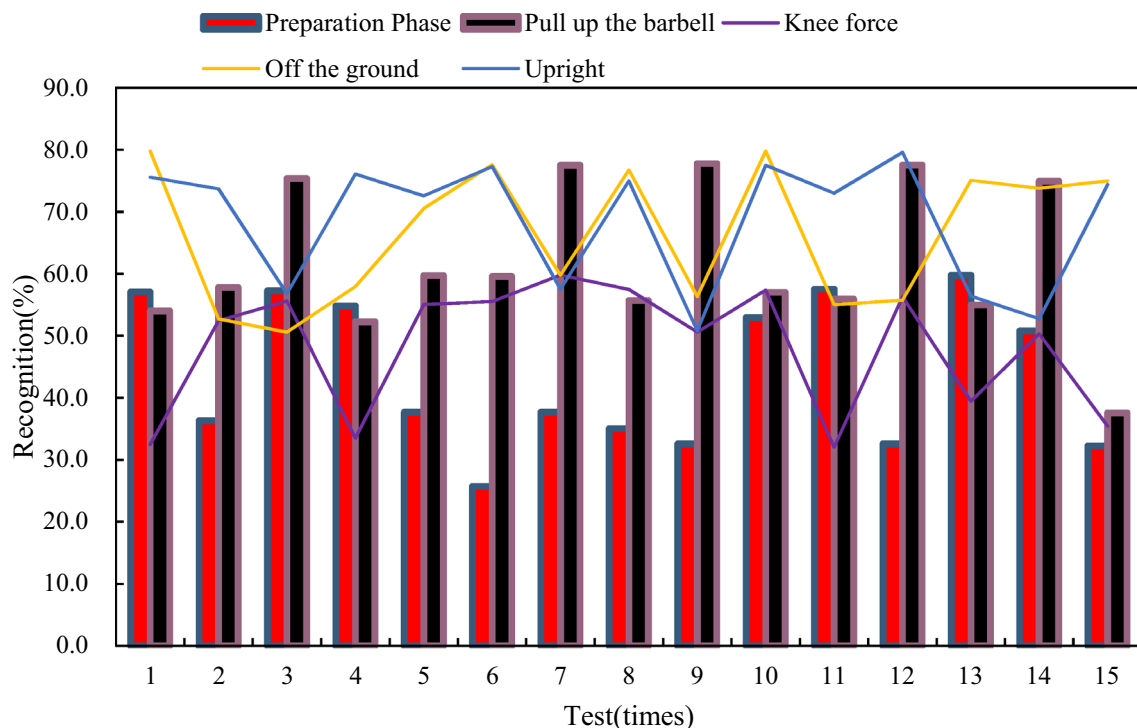


Fig. 2 The six-dimensional foot force information recognition of a single athlete

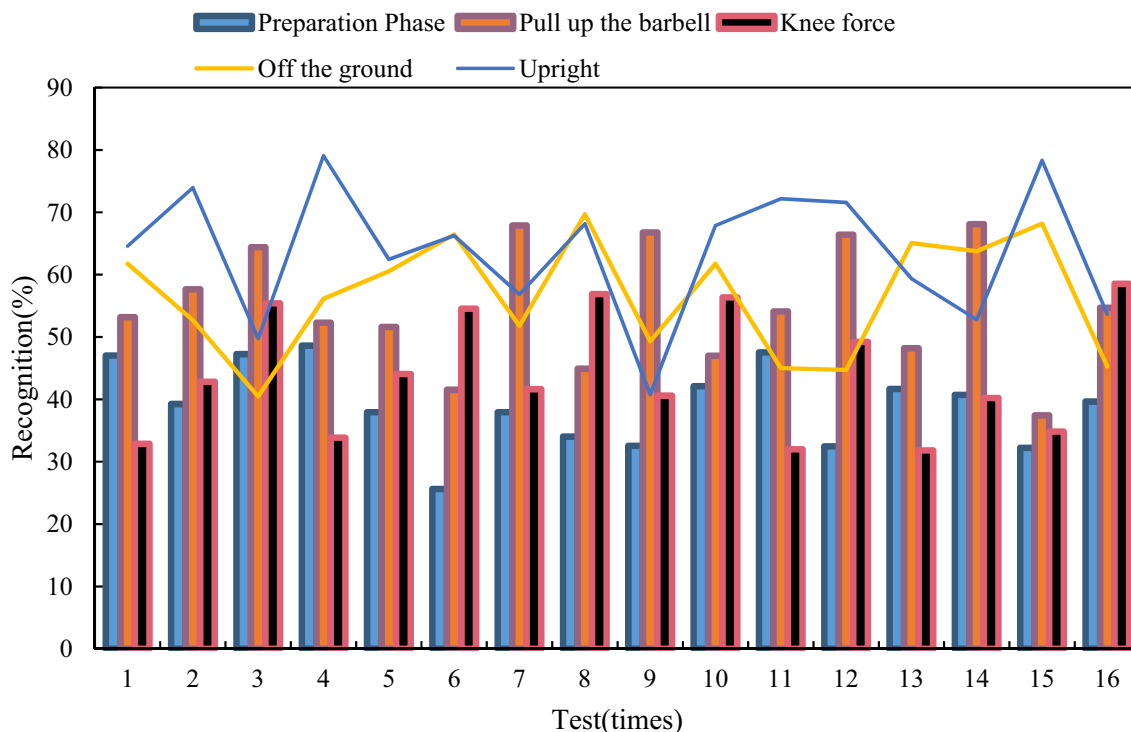


Fig. 3 Identification of six-dimensional foot strength information of multiple athletes

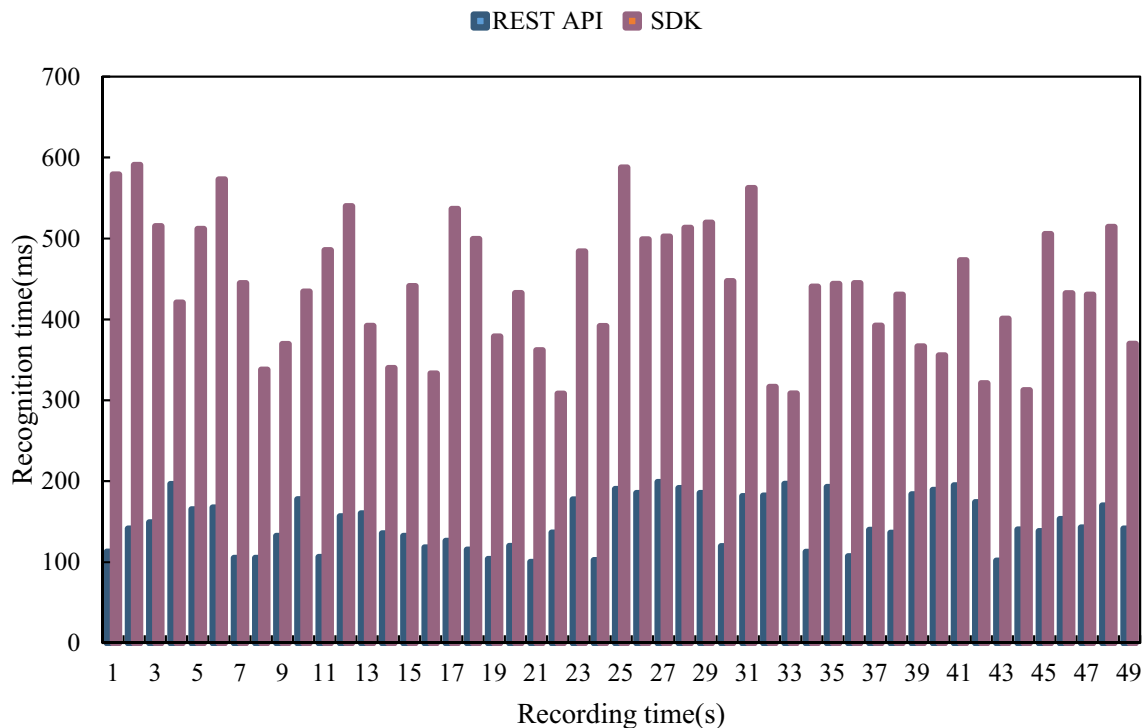


Fig. 4 Test results after REST API and Android SDK are integrated into Unity

characteristics of the rigid body connected with the joint and test the driving mode of the joint in the ODE simulation environment. Kinect can collect the spatial position

information of 20 key points of the human body. The human skeleton model is the basis of motion modeling. It specifies the calibration method of the human root

coordinate system and the tree relationship between the joint points. On the basis of the human skeleton model, the angle-axis description method is used to record the relative posture of each joint point. Finally, the motion data recorded by the angle-axis description method is converted into Euler angles to drive the virtual robot to simulate human motion.

The main reason why speech can express the emotional state of human beings is because some characteristic parameters in the speech signal can reflect the emotional state information. To identify the speaker's emotional state, we must first study which speech signal characteristic parameter differences will reflect changes in the emotional state. As a whole, the speech signal is a continuous process that changes with time, and it is not suitable for analysis using the processing technology of stationary signals. In order to reduce the truncation effect of the speech frame, the speech signal is framed and windowed. Although the voice signal changes continuously in time, it approximates a steady process in a short time, which is also called quasi-steady state. Therefore, in a short period of time, some physical characteristic parameters of speech, such as spectrum characteristics, remain basically unchanged, and the technology of processing stationary signals can be used to process short-term speech signals.

Human body movement is flexible and changeable, and it is difficult to use a unified mathematical model to describe it. In order to simulate human body movement, the form of motion data file is often used to save human body movement data. At present, common file formats for motion data include ASF/AMC, BVH, and C3D. The data structure saved in different formats of motion data files is slightly different, but the basic framework is similar, both of which are composed of human skeleton data and human motion data. When the human body moves, these sensors will record motion data; electromagnetic motion capture refers to the use of electromagnetic wave transmitters to emit electromagnetic fields in space according to a certain time and space law, and the receiving sensors installed on the biological body can receive electromagnetic signals and can receive received signal is transmitted to the data processing unit in real time to record the motion data; video-based motion data capture refers to the use of a camera to capture biological motion pictures, and analyze the captured organisms through the technology of image and video processing, computer vision, and pattern recognition. Motion data optical-based motion data capture refers to the use of image capture devices such as cameras to capture the motion of biological objects with optical markings. These optical marking points are generally located on key points such as joints, and then the positions of these nodes are recorded in real time, etc.

Computer simulation has always been an important research method in the field of robotics. In recent years, with the continuous development of computer simulation technology, a series of dynamic simulation software such as Adams, Unity3D, ODE, SolidWorks, etc. have appeared, and their characteristics and targeted application fields are also different. Adams is mainly aimed at the mechanical manufacturing industry and is used for statics, kinematics and dynamics analysis of mechanical systems. Unity3D is mainly used in the field of 3D games. It is a professional game engine. Because of its flexible interface and powerful physical simulation functions, many researchers also use it in the field of robot simulation.

Movement is a basic skill for simulating characters, but the design of the controller is particularly difficult. The control space is high-dimensional and the dynamics is nonlinear. It is very common that the controller is not feasible because the contact is unstable. Although robust walking controllers have been designed recently, the gait they produce looks unnatural. Whether it is a human or a computer, it is difficult to define control data suitable for natural walking.

5 Conclusions

The main research of this paper is to focus on the current situation that the existing wearable smart devices and their systems lack in-depth mining of the collected sports data and cannot accurately identify the types of human motions, resulting in a large amount of data resources being wasted seriously. Through sorting out the past related sports wearable intelligence Based on the related research results of the human motion recognition information processing system, a three-layer human motion recognition information processing system architecture is proposed. The system is mainly divided into three parts: data acquisition layer, data calculation layer, and data application layer. The data collection layer collects and preprocesses data; the data calculation layer is the core layer of the system, responsible for data storage, segmentation, feature extraction, and motion recognition classification; the data application layer is used to visually display recognition results and motion data. This paper applies the human motion data collected by the depth camera to the virtual character model of unity3d, integrates voice recognition third-party voice service platform, and realizes the conversion of voice-controlled virtual character 3D simulation animation. Experiments show that the method of controlling the virtual character's simulation movement through voice as the input interface is effective.

Authors' contributions JM Writing—editing data curation. JH Supervision and data analysis.

Funding The research was not specifically funded.

Availability of data and material Data sharing does not apply to this article because no data set was generated or analyzed during the current research period.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Consent for publication The picture materials quoted in this article have no copyright requirements, and the source has been indicated.

Ethics approval and consent to participate This article is ethical, and this research has been agreed.

References

- Ali J, Altintas B, Pulatkan A et al (2020) Open versus arthroscopic Latarjet procedure for the treatment of chronic anterior glenohumeral instability with glenoid bone loss. *Arthroscopy* 36(4):940–949
- Beibei, Hongliu, Qingyun et al (2019) [Study on gait symmetry based on simulation and evaluation system of prosthesis gait]. *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi* 36(6):924–929
- Carbone S, Moroder P, Runer A et al (2018) Scapular dyskinesis after Latarjet procedure. *J Shoulder Elbow Surg* 25(3):422–427
- Cerciello S, Redler A, Corona K (2018) Regarding, “Revision arthroscopic repair versus Latarjet procedure in patients with recurrent instability after initial repair attempt: a cost-effectiveness model.” *Arthroscopy* 34(4):1005–1006
- Chi Andrew S, Kim John et al (2017) Non-contrast diagnosis of adhesive capsulitis of the shoulder. *Clin Imaging* 7(44):46–50
- Cuéllar A, Cuéllar R, de Heredia PB (2017) Arthroscopic revision surgery for failure of open Latarjet technique. *Arthroscopy* 33(5):910–917
- Ekhtiari S, Horner NS, Bedi A et al (2018) The learning curve for the Latarjet procedure: a systematic review. *Orthop J Sports Med* 6(7):157–168
- Ernstbrunner L, Plachel F, Heuberger P et al (2018) Arthroscopic versus open iliac crest bone grafting in recurrent anterior shoulder instability with glenoid bone loss: a computed tomography—based quantitative assessment. *Arthroscopy* 34(2):352–359
- Fields BKK, Skalski MR, Patel DB et al (2019) Adhesive capsulitis: review of imaging findings, pathophysiology, clinical presentation, and treatment options. *Skeletal Radiol* 48(8):1171–1184
- Godenèche A, Merlini L, Roulet S et al (2020) Screw removal can resolve unexplained anterior pain without recurrence of shoulder instability after open Latarjet procedures. *Am J Sports Med* 48(6):1450–1455
- Harris G, Bou-Haidar P, Harris C (2019) Adhesive capsulitis: review of imaging and treatment. *Med Imaging Radiat Oncol* 57(6):633–643
- Homsí C, Bordalo-Rodrigues M, da Silva JJ et al (2020) Ultrasound in adhesive capsulitis of the shoulder: Is assessment of the coracohumeral ligament a valuable diagnostic tool? *Skeletal Radiol* 35(9):673–678
- Hsu Jason E, Anakwenze Okechukwu A, Warrender William J, Abboud JA (2019) Current review of adhesive capsulitis. *J Shoulder Elbow Surg* 20(3):502–514
- Hurley ET, Lim Fat D, Farrington SK et al (2019) Open versus arthroscopic Latarjet procedure for anterior shoulder instability: asystematic review and meta-analysis. *Am J Sports Med* 47(5):1248–1253
- Jurkovic IA, Papanikolaou N, Stathakis S et al (2016) Assessment of lung tumour motion and volume size dependencies using various evaluation measures. *J Med Imaging Radiat Sci* 47(1):30–42.e1
- Lafosse T, Amsallem L, Delgrande D et al (2017a) Arthroscopic screw removal after arthroscopic Latarjet procedure. *Arthrosc Tech* 6(3):559–566
- Lafosse L, Lejeune E, Bouchard A et al (2017b) The arthroscopic Latarjet procedure for the treatment of anterior shoulder instability. *Arthroscopy* 23(11):1–5
- Latarjet M (2020) Treatment of recurrent dislocation of the shoulder. *Lyon Chir* 49(8):994–997
- Lee SY, Park J, Song SW (2019) Correlation of MR arthrographic findings and range of shoulder motions in patients with frozen shoulder. *Am J Roentgenol* 198(1):173–179
- Lei F, Li X (2017) The kappa (κ) model of the Longmenshan region and its application to simulation of strong ground-motion by the Wenchuan MS8.0 earthquake. *Chin J Geophys—Chin Ed* 60(8):2935–2947
- Makhni EC, Lamba N, Swart E et al (2018) Revision arthroscopic repair versus Latarjet procedure in patients with recurrent instability after initial repair attempt: a cost-effectiveness model. *Arthroscopy* 32(9):1764–1770
- Owling PD, Akhtar MA, Liow RY (2016) What is a Bristow-Latarjet procedure? A review of the described operative techniques and outcomes. *Bone Joint J* 98(9):1208–1214
- Ranalletta M, Bertona A, Tanoira I et al (2018) Modified Latarjet procedure without capsule labral repair for failed previous operative stabilizations. *Arthrosc Tech* 7(7):711–716
- Siqueira JC, Pe Rh Inschi MG, Al-Sinbol G (2017) Simplified atmospheric model for UAV simulation and evaluation. *Int J Intell Unmanned Syst* 5(2–3):63–82
- Suh CH, Yun SJ, Jin W et al (2019) Systematic review and meta-analysis of magnetic resonance imaging features for diagnosis of adhesive capsulitis of the shoulder. *Eur Radiol* 29(2):566–577
- Wang W, Cao J, Zhang N et al (2017) Magnetic-spring based energy harvesting from human motions: design, modeling and experiments. *Energy Convers Manage* 132(Jan):189–197
- Xia S, Gao L, Lai YK et al (2017) A survey on human performance capture and animation. *J Comput Sci Technol* 32(3):536–554
- Xiang L, Pan Y, Gong C et al (2017) Adaptive Human-Robot Interaction Control for Robots Driven by Series Elastic Actuators. *IEEE Trans Rob* 33(1):169–182
- Yang C, Li Z, Cui R et al (2017) Neural network-based motion control of an underactuated wheeled inverted pendulum model. *IEEE Trans Neural Netw Learn Syst* 25(11):2004–2016
- Zhao W, Zheng X, Liu Y et al (2019) An study of symptomatic adhesive capsulitis. *PLoS ONE* 7(10):277–278
- Zhen-Bin G, Xiao-Li W, Jing SU et al (2018) Ecological compensation of Dongjiang river basin based on evaluation of ecosystem service value. *J Ecol Rural Environ* 34(6):563–570

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.