Scholar Publishing Group
International Journal of Educational Innovation and Science
https://doi.org/10.38007/IJEIS.2020.010104
ISSN 2789-7184 Vol. 1, Issue 1: 43-58

Computer Audio and Video Processing Technology in Sports Dance Teaching

Wenbo Cai*

Northwest Normal University, Gansu, China

caiwenbo@nwnu.edu.cn

**corresponding author*

Keywords: Audio and Video Processing Technology, Robot Technology, Sports Dance Teaching, Scientific Inquiry, Speech Recognition

Abstract: As society gradually becomes more intelligent, the development momentum of robots is becoming more and more fierce. Along with it is audio and video processing technology. In modern society, parents also pay more and more attention to the cultivation of children's hobbies; sports and dance are deeply loved by children. Therefore, sports dance robots combined with computer audio and video processing technology have gradually entered people's vision. This article mainly studies the influence of computer audio and video processing technology on the teaching of sports dance. In the experiment, 100 junior high school students were randomly selected. Through group teaching and group assessment and scoring, the comparative advantages of robot teaching and conventional teaching were explored. The experiment found that the students in the robot teaching group performed much better than the regular group. The lowest score in the robot teaching group was 60.9, while the lowest score in the conventional teaching group was 52.7; the robot teaching group scored 19 excellent people and 14 in the regular group. The results show that the robot teaching model has a certain role in promoting students' learning, and can well improve the students' enthusiasm for learning, and then improve their scores. Robots have a certain role in promoting modern classroom teaching, which can effectively improve students' enthusiasm for learning, and also greatly improve the learning efficiency of children. At the same time, audio and video processing technology has also been more widely used.

1. Introduction

1.1. Research Background and Significance

Robots have become more and more intelligent, hardware costs are getting lower and lower, and the application of robots has gradually expanded from restaurants and banks to supermarkets, schools and hospitals, and even the homes of every ordinary person. At present, the application of sports dance robots in the education industry is becoming more and more extensive. Parents are paying more and more attention to the development of children's art cells, and they have cultivated children's hobbies since childhood, and dance and sports are their only choices when training children, so this also promotes sports dance robots. The combination of sports dance robots and new computer audio and video processing technologies can stimulate children's interest in various ways, such as video playback, audio lectures, action demonstrations, and at the same time learn what they learn in a way that is easy for children to accept.

This article mainly studies the application of computer audio and video processing technology to dance movement teaching, so as to improve students' enthusiasm for learning, and promote the art of sports dance from monotonous to entertainment, from experience to comprehensive. Voice and video processing technology has brought a huge revolution to educational services and is also a challenge to voice and video technology itself.

1.2. Research Status and Existing Problems at Home and Abroad

The design and application of robots for sports dances abroad have been carried out for many years. At the same time, they have led the competition of many robots. Almost every year there are more than 100 robot competitions, and the participants are almost from all ages, including students [1]. Major overseas robot competitions include the International Robot Olympic Games, Robot World Championship and Robot World Cup [2]. In recent years, countries and governments have begun to recognize the importance of robot education, and actively introduced robot education into primary and secondary school classes [3]. The purpose is to make students interested in robots and show their creativity to develop their own personal robots. Many provinces and cities have incorporated robot education into the teaching guidelines, becoming one of the important indicators of teaching evaluation. Robot education has become an important indicator for evaluating students' creativity [4].

The continuous advancement of science and technology has enabled the rapid development of robotics, and educational robots have been integrated into our lives. Through analysis, it summarizes the six major trends in the development of educational robots: intelligence, popularity, diversity, openness, modularity, and networking [5-6]. With the development of robot technology, in recent years, remote telepresence robots have realized daily applications in many fields [7-8]. In the remote work area, the JazConnect robot uses an ultrasonic sensor and an infrared sensor to complete the fault avoidance function, uses a 30m laser sensor for autonomous navigation, and also sets up a video conference system. Users can access the client through a mobile phone or computer, and click on the picture to determine the moving direction and destination [9-10]. The robot can plan its own trajectory for movement. Within 2 meters of the fixed stopping point, it can automatically stop according to the instructions [11-12]. In the field of distance education, PEBBLES robots are equipped with a video conferencing system, while optimizing the teaching environment, submitting homework in real time, reading blackboards, and raising hands to answer questions [13-14].

At present, the video sources and display devices of video and audio processing systems are mainly analog signals. Before processing video data, digitization is necessary [15]. Digitization is done by a video decoder, and conversion of digital video to analog video is done by a video encoder [16-17]. With the emergence of many new audio and video applications, due to the lack of corresponding traffic, the defects of the intelligent recognition function are becoming increasingly apparent [18]. Therefore, in recent years, many cognitive methods of machine learning based on statistical features have emerged in the cognitive field of related research data [19]. In this type of method, the internal information characteristics of data packets are no longer the focus of research, but the macro characteristics of the overall flow within the network are used as research centers for identification [20]. In other words, the statistical characteristics of the business are used for classification and identification [21]. At the same time, it can also improve the adaptability of the recognition model, solve the problems of better face recognition, and improve the unknown traffic feature database and classification recognition [22]. Because the nature of the media stream is different, the QoS is also different [23]. When multiple media streams are multiplexed, the overall QoS evaluation standard of any media stream is used as a judgment to determine whether it is used. Generally, the highest overall QoS evaluation standard cannot be used as the standard [24]. This leads to inefficient use of network bandwidth and wastes a lot of rare bandwidth resources [25]. Second, audio data and video data cannot be deleted or added at will, nor can audio data and video data be processed separately [26]. Third, due to the limited application range of the method, when there are multiple sources, the method can only perform synchronization control in each source, and synchronization between sources is powerless [27].

1.3. Related Work

With the gradual deepening of robot technology, it has become a trend for robots to enter the classroom. Qin et al. believe that robot dance is an important theme in robot technology. The traditional robot dance system mainly relies on the rhythm or rhythm of music. However, the dance style of these conventional systems is limited and the novelty of the movements is poor. He instead developed a humanoid robot dancing system driven by musical structure and emotions. In his proposed system, music phrases and dance phrases are regarded as the basic structural units of music and dance, respectively. He created a phrase algorithm based on music theory, dividing a piece of music into a series of phrases. When the emotion of each phrase has been identified, an emotion sequence can be established. At the same time, the Hidden Markov Model (HMM) matches the dance phrase sequence with the emotion sequence. In particular, several concepts of the "opportunity method" proposed by choreographer Merce Cunningham are used to guide the robotic dance system. Therefore, he can arrange dance phrases by randomly selecting and combining multiple actions from a pre-designed action library [28]. Although his design framework is correct, there are many details. Jian studied the ant colony algorithm to improve the visual cognitive function of intelligent robots. Based on an in-depth understanding of the current status of research in this field at home and abroad, he draws on the achievements of cognitive science and neurobiology, and proposes solutions based on the structure and function of the human brain from the perspective of ant colony algorithms. He simulated an autonomous learning process controlled by human long-term memory and its working memory, and proposed an autonomous learning algorithm for long-term memory growth driven by visual strangeness. His method takes incremental self-organizing network as a long-term memory structure, and combines Q-learning method with visual unfamiliar intrinsic motivation for working memory. The visual knowledge obtained through self-learning is continuously accumulated into long-term memory, thereby achieving the ability of self-learning, memory and intellectual development similar to humans [29]. Although his research is more innovative, it lacks accuracy. Sha believes that in the past decade, due to the widespread deployment of athletes and ball tracking systems in team sports, the analysis in professional sports has experienced a phenomenal increase. For example basketball and football. With the generation of large amounts of fine-grained data, new data points are also being generated, which can clarify the performance of players and teams. However, due to the complexity of competition in continuous sports, these data points often lack the specificity and environment for meaningful retrieval and analysis. He proposed an intelligent human-machine interface that uses trajectories instead of words, so that specific game content can be retrieved in motion. His system uses a variety of alignment, templating and hashing technologies, which are tailored for multi-agent scenarios, enabling interactive speed [30]. Although his method is more detailed, the amount of data is huge and difficult to analyze. Sengupta believes that consumer electronics (CE) systems are flooding our daily lives, from smartphones, smart devices, set-top boxes, video consoles; tablet PCs to washing machines and kitchen appliances. He believes that the successful intelligent functions of CE equipment depend heavily on audio and video technologies. With the development of these complex devices and technologies, the security of data (such as images, sensitive information, etc.) has become a major issue that requires immediate attention. He believes that another aspect that needs attention in these technologies is to optimize the data redundancy, coverage, and energy consumption of CE systems (such as video equipment) [31]. Although his idea is correct, it is not novel and lacks research data as a supporting point.

1.4. Innovation

In this paper, through the study of computer audio and video processing technology and robot technology, it is used in modern classroom teaching to explore the impact of sports dance robots on student learning, and the use of comparative experiments proves that sports dance robots can improve students' enthusiasm for learning. Students' learning efficiency also has a positive impact, and can inspire students' love of technology.

2. Audio and Video Processing Technology

2.1. Network Traffic Identification Technology

Network flow identification technology mainly includes three parts: network flow identification technology based on ports, deep packets and statistical functions. Its main architecture is shown in Figure 1.

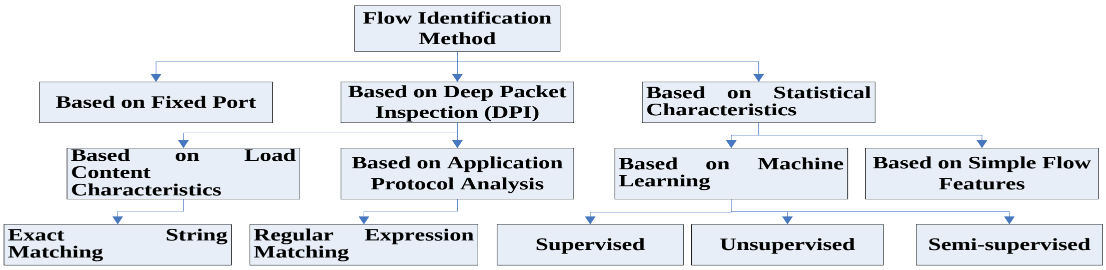

Figure 1. Main architecture of network traffic identification technology

(1) Port-based network traffic identification technology

This technology is usually based on the 16-bit port number of the network transport layer to identify different network protocols and service processes, so as to further achieve the purpose of traffic classification. This method is usually not affected by the password, and can complete the classification operation quickly and accurately.

(2) Network traffic identification technology based on deep packet inspection

DPI technology deeply analyzes the payload portion of each packet of the network flow to match a given set of patterns to identify specific application categories and protocol definitions. This method solves the misjudgment of using dynamic or pseudo ports without considering the network port number. According to different implementation methods, this method is divided into a recognition method based on load content feature matching and a recognition method based on application protocol analysis. The identification method based on the characteristics of the payload content is to detect the depth of the packet payload, read the signature of the signature, and identify the corresponding application type.

(3) Network traffic identification technology based on statistical characteristics

Most statistical feature recognition methods use the ML method. The ML method requires training and learning on the training set, using the validation set to verify the classification effect of the classifier, and finally identifying and classifying the unknown sample set. The ML-based network flow classification method can enhance the classification effect of flow recognition, because a better feature subset can be selected through a feature selection algorithm.

2.2. Mixing Technology

The structure model of the mixer is shown in Figure 2. In the actual mixing process, it is impossible to put the user's voice into the mixer during mixing, to avoid hearing the echo in the repeated sound after mixing, that is, $a_j^i(t)$ outputs corresponding echo $b_j^i(t)$ to the user's repeated mixer, and does not contain the $a_j^i(t)$ related components.

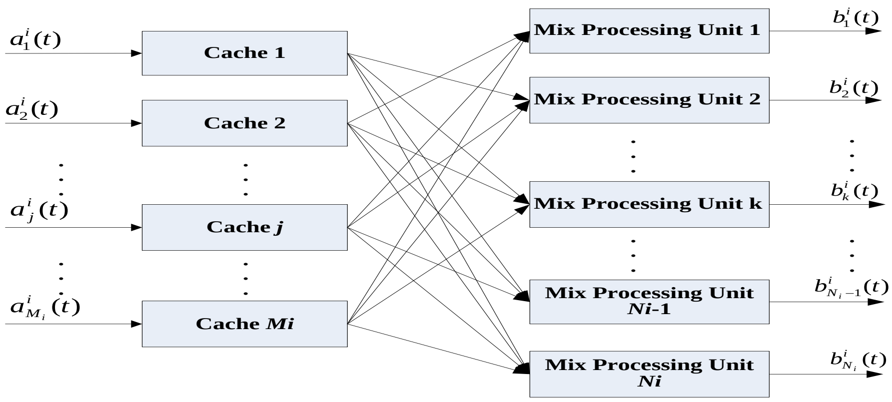

Figure 2. Structural model of the mixer

Each mixer has multiple mixer processing units, which is an effective algorithm logic unit, which can improve the performance of the mixer in the audio processor. The mixer processing unit can perform mathematical operations on multiple input digitized audio signals and convert them into outputs. Generally, the mathematical relationship of the mixer processing unit processing multiple audio signals can be expressed as:

$$\begin{cases} b_k(t) = \sum_{j=1, j\neq k}^{M_i} a_j^i(t), i = 1,2,...,K; k = 1,2,...,M_i \\ b_{N_{i+1}}(t) = \sum_{j=1}^{M_i} a_j^i(t), i = 1,2,...,K \end{cases} \tag{1}$$

In the formula, $b_k(t)$ is the audio data output from the inside after mixing, so it is not involved in mixing. $b_{N_{i+1}}(t)$ is the audio data for all users to participate in the mixing process.

Suppose that at time t, the data $a_j^i(t)$ after speech decoding in the j-th path of the i-th meeting entity has the corresponding weight $w_j^i(t)$, then the general weight calculation formula of the mixer algorithm is as follows.

$$\begin{cases} b_k^i(t) = \sum_{j=1, j\neq k}^{M_i} w_j^i(t) a_j^i(t) / \sum_{j=1, j\neq k}^{M_i} w_j^i(t), i = 1,2,..., K; k = 1,2,..., M_i \\ b_{N_{i+1}}^i(t) = \sum_{j=1}^{M_i} w_j^i(t) a_j^i(t) / \sum_{j=1}^{M_i} w_j^i(t) \end{cases} \quad (2)$$

(1) Average adjustment weight method

The average weight adjustment method is to average the weights of audio signals involved in the mixing process, that is, the weights have the same value. Take $w_j^i(t) = \frac{1}{M_i}$, then:

$$\begin{cases} b_k^i(t) = \frac{1}{M_i - 1} \sum_{j=1, j\neq k}^{M_i} a_j^i(t), i = 1,2,..., K; k = 1,2,..., M_i \\ b_{N_{i+1}}^i(t) = \frac{1}{M_i} \sum_{j=1}^{M_i} a_j^i(t) \end{cases} \quad (3)$$

(2) Strong alignment weight method

This method mainly designs weighting function based on the amplitude value of signal involved in mixing processing. It is specified that in the mixer processing unit of the ith meeting entity at time t, the maximum value of the data in the buffer module of all the input participating in the mixing is $BuMax^i$, and the maximum value of $b_k^i(t)$ can be calculated as $MixMax^i$ through calculation, then:

$$\hat{b}_k^i(t) = b_k^i(t) \cdot \frac{BuMax^i}{MixMax^i} \cdot \mu^i \quad (4)$$

Here, the size of the change value μ^i is adjusted as the adjustment coefficient $\hat{b}_k^i(t)$. The range of μ^i is the area adjacent to $\left|\frac{BuMax^i}{MixMax^i}\right|$. By readjusting the buffer length of the mixing processing unit, the mixing result $\hat{b}_k^i(t)$ is adjusted to a smaller range.

(3) Real-time weighting method

The basic idea of the real-time weighting method is to weight the ratio of multiple voice signals participating in the mixing process, then:

$$w_j^i(t) = \frac{|a_j^i(t)|}{\sum_{p=1}^{M_i} |a_p^i(t)|} \quad (5)$$

2.3. MPEG-2 Standard System Structure

Digital signals such as audio and video streams are compressed and encoded to form their respective elementary code streams (ES), but these elementary code streams cannot be stored or transmitted directly, so they must be sent to the packer. In this way, it can be subdivided according to a certain format and specific identification characters. The audio/video stream needs to be decoded accurately. The same PES package only contains video or audio data, because the purpose of packaging is to package audio and video signals respectively. The audio access unit is an audio frame, the video access unit is an image frame, and the length of the PES packet is usually an access unit.

After the PES packets are formed by each wrapper, the audio/video stream data will be sent to two subsystems, program stream (PS) and transmission stream (TS). The transport stream subsystem time-division multiplexes one or more PES data streams and combines them into a single code stream. These PES packets have several independent time benchmarks and also a common time benchmark. Because of the fixed length of data packets in the transmission stream, the decoder can easily find the synchronization information, and once the synchronization information is lost, the system can recover, so the transmission stream is used in a relatively harsh channel environment, such as cable, satellite or radio video transmission medium. The program stream subsystem uses time-division multiplexing to decode the PES data stream synchronously through one or more common time references. For example, the audio or video code stream of the program can be combined with one code stream. When the program stream synchronization is lost, it may be difficult for the decoder to re decode the synchronization information because the packet length of the program stream is not fixed.

2.4. Silent Detection Technology

Silent detection algorithms include short-time energy detection, short-time zero-rate detection, spectral feature detection, spectral entropy detection, etc. Among them, silent detection algorithms based on energy features and spectral entropy features are very common in VoIP systems.

(1) Short-term energy method

For the speech waveform time domain signal $x(n)$, the short-term energy is defined as:

$$En = \sum_{m=-\infty}^{\infty} [x(m) \cdot w(n-m)]^2 \tag{6}$$

In the formula, when the frame division process is performed as a window function $w(m)$, the audio signal $x_k(m)$ of the k frame satisfies the following expression:

$$x_k(m) = w(m)x(k+m), 0 \le m \le N-1 \tag{7}$$

In the formula, $k = 0, 1T, 2T, \ldots$, N is the length of the frame, and T is the length of the frame movement.

The short-term energy E_k is:

$$E_k = \sum_{m=k-(N-1)}^{k} [x(m) \cdot w(k-m)]^2 \tag{8}$$

As can be seen from the above equation, the short-term energy is actually a weighted square sum of the sample values of one frame of audio signal. Therefore, the short-term energy varies significantly with the amplitude of the audio signal, especially for high-level audio signals.

(2) Spectral entropy algorithm

The spectrum entropy algorithm studies the frequency spectrum of the speech signal in the frequency domain. In a wide-band speech signal, the power spectrum of the effective speech signal has a more concentrated distribution than the entire spectrum area, and the fluctuation is large, so the spectral entropy as the average amount of information is small. Since the noise signal has a relatively flat distribution in the entire spectrum area, its average information amount, that is, the spectrum entropy value is large. The construction process of the spectral entropy algorithm is shown in Figure 3.

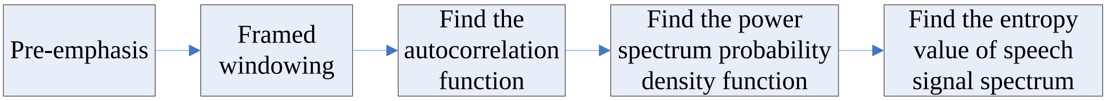

Figure 3. Construction process of spectral entropy algorithm

The specific process of the algorithm is as follows:

1) First pre-emphasize the voice signal $s(n)$.

2) By adding a window function $w(n)$ and then framing the windowed audio signal $s(n)$.

$$s_w(n) = s(n) \cdot w(n); n = 0,1,\ldots, N-1 \tag{9}$$

3) The short-term autocorrelation function of each frame is:

$$r(n) = \sum_{k=0}^{N-n-1} s_w(k) s_w(k+n); n = 0,1,\ldots, N-1 \tag{10}$$

4) Perform FFT transformation to obtain the short-term power spectral density function:

$$s_w(k,m) = \sum_{n=0}^{N-1} r(n) e^{-j\frac{2\pi kn}{N}}; k = 0,1,\ldots, N-1 \tag{11}$$

5) For the k-th frequency component in frame m, the normalized PSD function is:

$$P(k,m) = \frac{s_w(k,m)}{\sum_{k=0}^{N-1} s_w(k,m)}; k = 0,1,\ldots, N-1 \tag{12}$$

6) The calculation formula of power spectral entropy of frame m is as follows:

$$H(m) = \sum_{k=0}^{\frac{N}{2}-1} P(k,m) \log\left[\frac{1}{P(k,m)}\right] = -\sum_{k=0}^{\frac{N}{2}-1} P(k,m) \log P(k,m) \tag{13}$$

The quantization level is essentially the quantization value of the nominal amplitude, usually an integral power of 2. The higher the quantization level and accuracy, the closer the analog signal reconstructed by digital signal is to the actual signal. The quantization bits are actually the bits that store a sample value in the computer, usually 8, 12 and 16 bits. There is a correlation between quantization bits and quantization level. The more quantization bits, the larger quantization level, and the smaller quantization error, the smaller signal distortion, but the larger the occupied space.

3. Robot Sports Dance Teaching

3.1.Limb Technology Movement Synthesis Technology

The teaching model of robot sports dance is shown in Figure 4.Using the motion path editing in the toolset layer, the modular dance designer can specify a new motion path for further designed motions, mainly at this stage for collision detection and path avoidance between virtual humans, and re-edit the considered route. After completing the above tasks, the dance designer can preview the designed actions and further modify them. When the dance designer completes the motion design for each stage, it can save the motion and design the next stage, or leave the system. In order to ensure a smooth transition between the designed multi-step motions, the system will automatically perform transitional motion synthesis between motion segments corresponding to adjacent segments. These designated "connection actions" are fed back to the dance designer as "candidate connection actions", and the dance designer selects appropriate actions from the "candidate connection actions" to add to the system. Then, use the "motion connection" of the tool layer to connect the motion in segments, and the overall motion will be rough. The next step is to use "motion mixing" to smooth the motion, and finally use "machine motion synthesis" to add equipment. By using the three modules in the order in which they work together, you will be able to get dance clips that satisfy the dance designer.

Figure 4. Teaching model of sports dance robot

3.2. Sports Editing

The generation of new action sequences includes three steps:

(1) The target operation request includes basic information, such as the target operation type and length.

(2) Organize the movement and combine the selected movement fragments with the movement fragments captured by the movement.

(3) Smooth transition, processing actions that do not meet the mechanical requirements of the final synthetic animation.

The pose f_i, f_j of any two pose data, the distance of the center of gravity:

$$DR(f_i, f_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (14)$$

In the formula, (x, y, z) represent the coordinates of the center of gravity of the person.

The similarity of human posture directions is:

$$Dd(fi, fj) = \sum_{k=1}^{S} \left\| J_i - J_i' \right\|^2 \quad (15)$$

In the formula, J is a measure of the overall freedom of the human body.

Let f_{ij} be the transport distance from position A to B, and d_{ij} be the ground distance between AB. The distance of Mont-Kantlovich is:

$$\text{Monge-Kantorovich} = \frac{\min_F \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \quad (16)$$

In practical applications, frame insertion is performed based on the high point, low point, inflection point, etc. of an action, or the intersection of two or more actions. The frame at this time can be regarded as a starting point, and then look for it the next frame node.

4. Comparison Experiment between Robot Teaching Group and Conventional Teaching Group

4.1. Experimental Environment and Data Set

The operating environment and experimental data set of this experiment are shown in Table 1 and Table 2.

Table 1. Operating environment

Project	Category
CPU	Intel(R)Xeon(R)CPUE5-2630v3@2.40GHz
RAM	8G
System	Win7
Development tools	MatlabR2016a
Data collection	Wireshark
Data processing	Shell script

Table 2. Experimental data set

Business type name	Sample bytes (Byte)
ASD	0.61G
AHD	1.24G
HTTP download	2.86G
Sopcast	2.31G

4.2. Subject

Randomly select 100 junior high school students and divide them into two groups: the robot teaching group and the regular teaching group. The former is taught in the form of sports dance robot teaching, while the latter is taught in the form of regular teacher face-to-face teaching.

4.3. Experimental Process

(1) Divide the difficult dance in one section into three classes, and the duration of each class is 1 hour, to ensure that the content of the two groups of the experiment is the same.

(2) The robot teaching group records the learned gymnastics and dance into the robot brain chip, and interacts with students through voice explanation and video playback functions. The teaching mode of the robot teaching group is shown in Figure 5. The instructional video is imported into the robot brain, and the robot preprocesses the data, then calculates the difficulty level of the action, and further classifies the actions according to the difficulty level, thereby completing the teaching task. Robot teaching first plays the course in the form of a video, and then records the recorded manual explanation. After that, the robot will demonstrate the movement and decompose the complex movement to speed up the students' understanding.

(3) The general teaching group is to prepare lessons first, and then the teacher takes 1 hour to teach the students.

(4) After the completion of the three teaching tasks, a professional teacher will conduct a unified inspection. The inspection results are divided into excellent, good, and poor. The requirements are greater than or equal to 85 to be excellent, and between 65 and 85 points are good. If it is less than 65, the score is poor, and the result will be recorded.

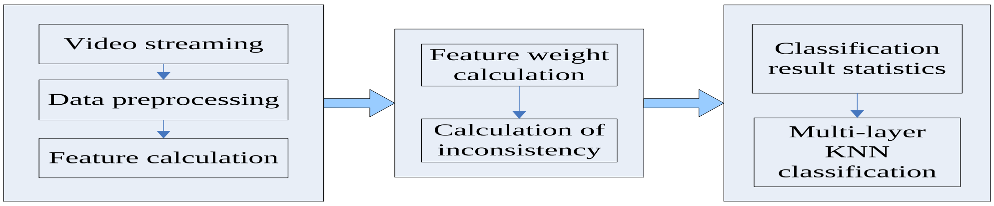

Figure 5. Block diagram of robot teaching mode

4.4. Data Analysis

All data were statistically analyzed according to the robot teaching group and the conventional teaching group, and the excellent and poor performance curves of each group were obtained.

5. Analysis of the Application of Sports Dance Robots in Teaching

5.1. Analysis of Comparative Experiment Results

Table 3. Some experimental results

Robot teaching group	80.2	91.9	82.3	76.7	81.7
	88.1	91.2	81.6	77.8	77.5
	85.2	97.0	92.5	87.1	89.5
	62.7	84.2	94.8	88.8	91.3
	62.8	88.5	97.4	85.7	87.1
	60.9	87.1	94.3	99.2	99.2
Routine teaching group	86.2	75.7	81.7	85.1	76.9
	86.1	92.8	98.8	93.5	94.7
	82.8	52.7	64.7	55.2	85.8
	83.9	56.5	64.6	53.9	86.5
	81.3	61.7	57.3	63.5	84.5
	89.6	86.8	87.4	86.6	86.8

Through the examination of the teaching achievements of 100 junior high school students, it was found that the results of the robot teaching group were far better than those of the conventional teaching group. Some experimental results are shown in Table 3. It can be seen from the data in the table that the results of the students in the robot teaching group are much better than those in the regular group. The lowest score in the robot teaching group was 60.9, while the lowest score in the conventional teaching group was 52.7; the robot teaching group scored 19 excellent people and 14 in the regular group. From this, it can be judged that the robot learning model has a certain role in promoting students' learning, which can well improve students' interest in learning, stimulate their curiosity, and further improve their scores.

The statistical results of 100 junior high school students are shown in Table 4. It can be seen intuitively from the table that there are 29 people in the robot teaching group with a score higher than 85, which is 10 higher than the regular group; the number of students in the robot teaching group between 65 and 85 is 15, which is higher than that in the regular group 6 fewer; the number of students with less than 65 points in the regular group is 4 higher than that of the robot teaching group. From this, it can be seen that the robot teaching group can play an active role in the teaching of sports dance, making students essentially want Go to study and eliminate the repulsion of learning.

Table 4. Two sets of performance statistics

Grouping \ Score situation	⩾85	65⩽Score<85	<65
Robot teaching group	29	15	6
Routine teaching group	19	21	10

5.2. Analysis of Students' Acceptance of Robots

After the assessment, 60 students were randomly selected for a questionnaire survey, and their acceptance of the robot teaching mode was counted. The statistical results are shown in Figure 6. It can be seen from the figure that for this new teaching mode, nearly 50% of students are very fond of this new learning method and have a strong learning interest. It can be seen that students are very recognized for the form of robot teaching, and are willing to accept the combination of audio and video processing technology and robots appearing in modern classrooms.

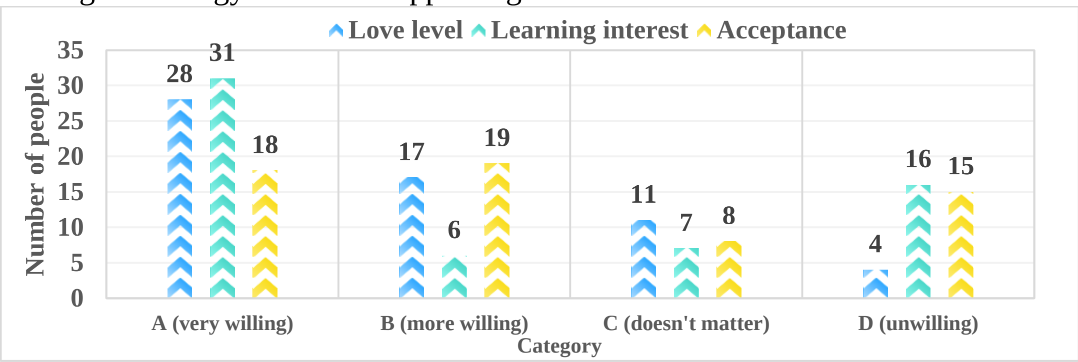

Figure 6. Survey results statistics

5.3. Analysis of the Memory Occupied by the Robot Brain during Operation

Figure 7 shows the robot memory usage before and after grouping. It can be seen from the figure that the memory consumed when running before grouping is very high, reaching a maximum of 345MB; after grouping, the memory consumption is significantly reduced, and the maximum memory size is only 153MB. After combining audio and video processing technology, the minimum memory occupied by the robot when running is 23MB. It can be seen that the grouped robot DFA engine has a much smaller memory footprint than the single DFA engine before grouping, and the memory before grouping follows the regular expression the double-speed growth of the length and wildcards increased sharply, but the memory growth rate slowed down after grouping.

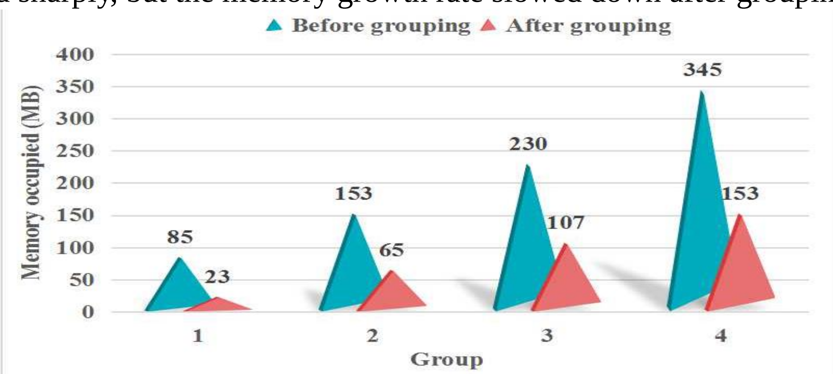

Figure 7. Comparison of memory usage before and after grouping

5.4. Comparative Analysis of Audio and Video Processing Technology before and after Application

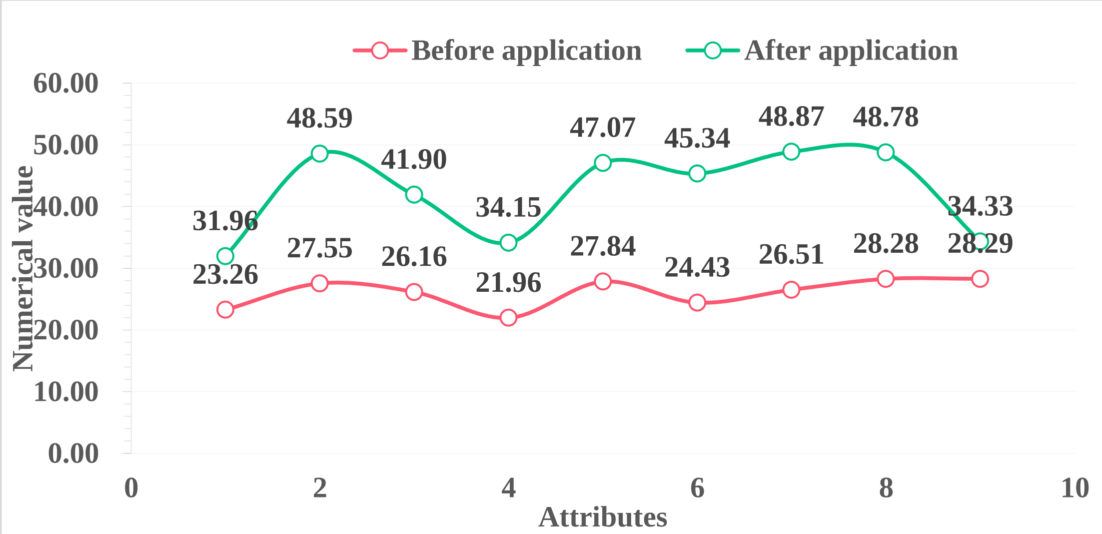

Figure 8. Before and after application of audio and video processing technology

The comparison of audio and video processing technology before and after the application of sports dance teaching is shown in Figure 8. As can be seen from the figure, when the audio and video processing technology is not used, the enthusiasm of the students is relatively low, and the range of change is relatively small; after applying it in teaching, the enthusiasm of the students has been significantly improved, the maximum value is close to 50. Therefore, audio and video processing technology has a promoting effect on the teaching of sports dance, and the effect is more obvious. If the audio cache does not overflow and does not disappear, the size design of the audio cache means that it can completely compensate for the jitter of the transmission delay. Therefore, media synchronization is also solved.

5.5. PSNR Value Analysis of Synthesized Sequences

The PSNR value of the actual synthesized sequence generated by the non-offset synthesis method and the pixel domain synthesis method relative to the ideal synthesized sequence is shown in Figure 9. It can be seen from the figure that the actual synthesized sequence has high similarity to the ideal high PSNR meaning synthesis, which shows the quality and availability of the sequence synthesis sequence, compared with the PSNR value of the sequence synthesis method without offset image synthesis, The PSNR value of the pixel domain synthesis method is lower, indicating that the actual synthesis effect of the non-offset video synthesis method is better than that of the pixel domain synthesis method. The bit rate of the synthesized stream tested is slightly higher than the normal bit rate, and the reason for the increase in the bit rate is that the synthesis source residuals tend to have higher energy in the corresponding input bit stream. This is because the prediction parameters reused from the input bitstream may not be the optimal parameters for synthesizing the bitstream. The offset-free video synthesis method is realized by maintaining the ideal synthesis sequence and the actual synthesis sequence in each step of the synthesis process. The method of compressing the area and the area of the pixels are used in combination so as not to cause any offset between the input video sequence and the composite video sequence. This method first decodes the input video stream, then reuses important encoding parameters (for example, encoding mode and motion vector) in the input stream, and partially re-encodes to form a composite video stream.

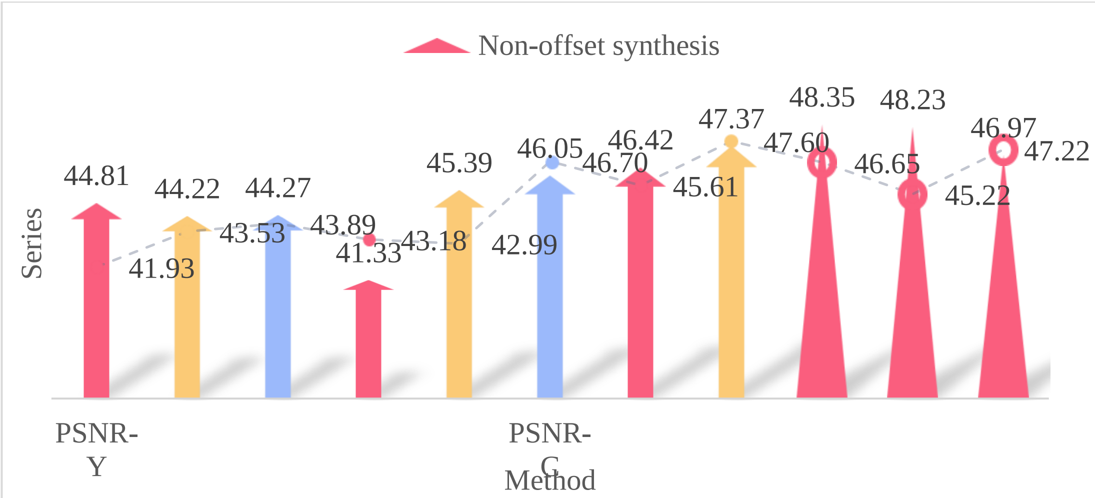

Figure 9. PSNR value of synthesized sequence

6. Conclusion

This article is mainly about the research of computer audio and video processing technology in sports dance teaching. Through the application of sports dance robots in actual teaching, the impact of the combination of robot technology and audio and video processing technology on sports dance teaching is explored. The study found that robots have a certain role in promoting modern classroom teaching, can effectively improve students' learning enthusiasm, and also greatly improve the learning efficiency of children.

The rapid development of audio analysis theory makes the task of designing audio and video synchronization systems in the music-driven mode more urgent. The integration of audio and video not only provides a highly practical art platform for the realization of music-driven editing concepts, but also promotes the innovation of sports and dance through the use of advanced technology.

Applying audio-visual processing technology to robot research, combined with sports dance education, can ensure the realism and fluency of the synthesized video, and also avoid the simple use of animation jumping and linear interpolation. Robots will be able to move their bodies and demonstrate accurate movements. Then, the difficult movements can be broken down so that students can learn all movements quickly. The scientific inquiry education in the robot course improves the positive attitude of robot learning and the positive attitude towards scientific inquiry, and helps students develop their ability to analyze and solve problems autonomously.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Wang Q, Yang X, Huang Z, et al. A novel design framework for smart operating robot in power system. IEEE/CAA Journal of Automatica Sinica, 2018, 5(2):531-538. https://doi.org/10.1109/JAS.2017.7510838

[2] Guanci Y, Jing Y, Weihua S, et al. Convolutional Neural Network-Based Embarrassing Situation Detection under Camera for Social Robot in Smart Homes. Sensors, 2018, 18(5):1530-. https://doi.org/10.3390/s18051530

[3] Chen Y, Wang W, Biron Z, et al. A Smart Companion Robot for Heavy Payload Transport and Manipulation in Automotive Assembly. IEEE Robotics & Automation Magazine, 2018, PP(99):1-1.

[4] Kim D G. Smart construction robot technology to improve construction and safety in outer walls of high-rise buildings. Asia life sciences, 2018(4):2303-2318.

[5] Wilson G, Pereyda C, Raghunath N, et al. Robot-enabled support of daily activities in smart home environments. Cognitive Systems Research, 2019, 54(MAY):258-272. https://doi.org/10.1016/j.cogsys.2018.10.032

[6] Masamune K, Nishikawa A, Kawai T, et al. The development of Smart Cyber Operating Theater (SCOT), an innovative medical robot architecture that can allow surgeons to freely select and connect master and slave telesurgical robots. Impact, 2018, 2018(3):35-37. https://doi.org/10.21820/23987073.2018.3.35

[7] Parashar S, Tomar P. Waste Management by a Robot- A Smart and Autonomous Technique. Iosr Journal of Electronics & Communication Engineering, 2018, 13(6):31-36.

[8] Duckworth P. Unsupervised Human Activity Analysis for Intelligent Mobile Robot. Ki Künstliche Intelligenz, 2018, 32(4):291-294. https://doi.org/10.1007/s13218-018-0558-4

[9] Gu S, Zeng W, Jia Y, et al. Intelligent Tennis Robot Based on a Deep Neural Network. Applied Sciences, 2019, 9(18):3746. https://doi.org/10.3390/app9183746

[10] Xiao-Wei X. Study on the intelligent system of sports culture centers by combining machine learning with big data. Personal and Ubiquitous Computing, 2020, 24(1):151-163. https://doi.org/10.1007/s00779-019-01307-z

[11] Gong W, Tong L, Huang W, et al. The optimization of intelligent long-distance multimedia sports teaching system for IOT. Cognitive Systems Research, 2018, 52(DEC.):678-684. https://doi.org/10.1016/j.cogsys.2018.08.005

[12] Subramaniyaswamy V, Logesh R, Indragandhi V. Intelligent sports commentary recommendation system for individual cricket players. International journal of advanced intelligence paradigms, 2018, 10(1-2):103-117. https://doi.org/10.1504/IJAIP.2018.089492

[13] Cao S. Intelligent lighting control system in large-scale sports competition venues. Light and Engineering, 2018, 26(4):172-182. https://doi.org/10.33383/2018-127

[14] Gu D, Lima S, Rocha, Álvaro. Analysis of tactical information collection in sports competition based on the intelligent prompt automatic completion algorithm. Journal of Intelligent & Fuzzy Systems, 2018, 35(3):2927-2936. https://doi.org/10.3233/JIFS-169648

[15] Munz M, Engleder T. Intelligent Assistant System for the Automatic Assessment of Fall Processes in Sports Climbing for Injury Prevention based on Inertial Sensor Data. Current Directions in Biomedical Engineering, 2019, 5(1):183-186. https://doi.org/10.1515/cdbme-2019-0047

[16] Kaiyan H, Qin W. The development and popularisation of network platform of college sports venues in intelligent manufacturing. International Journal of Reasoning-based Intelligent Systems, 2018, 10(1):20-25. https://doi.org/10.1504/IJRIS.2018.091126

[17] Guangjing L, Cuiping Z. Research on static image recognition of sports based on machine

learning. Journal of Intelligent and Fuzzy Systems, 2019, 37(12):1-11. https://doi.org/10.3233/JIFS-179203

[18] Liang S. Intelligent decision model of sports training knowledge based on dynamic deep learning. IPPTA: Quarterly Journal of Indian Pulp and Paper Technical Association, 2018, 30(6):174-181.

[19] Han K, Wang Q. Research on 020 Platform and Promotion Algorithm of Sports Venues Based on Deep Learning Technique. International Journal of Information Technology & Web Engineering, 2018, 13(3):73-84. https://doi.org/10.4018/IJITWE.2018070105

[20] Simoens P, Dragone M, Saffiotti A. The Internet of Robotic Things. International Journal of Advanced Robotic Systems, 2018, 15(1):172988141875942. https://doi.org/10.1177/1729881418759424

[21] Khosiawan Y, Khalfay A, Nielsen I. Scheduling unmanned aerial vehicle and automated guided vehicle operations in an indoor manufacturing environment using differential evolution-fused particle swarm optimization. International Journal of Advanced Robotic Systems, 2018, 15(1):172988141775414. https://doi.org/10.1177/1729881417754145

[22] Dekan M, Frantiek, Duchoň, Andrej B, et al. Moving obstacles detection based on laser range finder measurements. International Journal of Advanced Robotic Systems, 2018, 15(1):172988141774813. https://doi.org/10.1177/1729881417748132

[23] Maximo C, Sandra R. SART3D: A MATLAB toolbox for spatial audio and signal processing education. Computer Applications in Engineering Education, 2019, 27(4):971–985. https://doi.org/10.1002/cae.7

[24] Chu N N. The ieee international conference on video and audio signal processing in the context of neurotechnology [conference reports]. Consumer Electronics Magazine IEEE, 2019, 8(2):8-9. https://doi.org/10.1109/MCE.2018.2880847

[25] Gannot S, Naylor P A. Highlights From the Audio and Acoustic Signal Processing Technical Committee [In the Spotlight]. IEEE Signal Processing Magazine, 2019, 36(2):136-134. https://doi.org/10.1109/MSP.2018.2888634

[26] Li H, Zhu J, Ma C, et al. Read, Watch, Listen, and Summarize: Multi-Modal Summarization for Asynchronous Text, Image, Audio and Video. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(5):996-1009. https://doi.org/10.1109/TKDE.2018.2848260

[27] Ruan F, Zhang X, Zhu D, et al. Deep learning for real-time image steganalysis: a survey. Journal of Real Time Image Processing, 2020, 17(1):149-160. https://doi.org/10.1007/s11554-019-00915-5

[28] Qin R, Zhou C, Zhu H, et al. A Music-Driven Dance System of Humanoid Robots. International journal of humanoid robotics, 2018, 15(5):1850023.1-1850023.31. https://doi.org/10.1142/S0219843618500238

[29] Jian Y, Li Y. Research on intelligent cognitive function enhancement of intelligent robot based on ant colony algorithm. Cognitive Systems Research, 2019, 56(AUG.):203-212. https://doi.org/10.1016/j.cogsys.2018.12.014

[30] Sha L, Lucey P, Yue Y, et al. Interactive Sports Analytics: An Intelligent Interface for Utilizing Trajectories for Interactive Sports Play Retrieval and Analytics. Acm Transactions on Computer Human Interaction, 2018, 25(2):13.1-13.32. https://doi.org/10.1145/3185596

[31] Sengupta A, Ray N. Audio and Video Technologies: Recent Advances in Consumer Electronics. Consumer Electronics Magazine, IEEE, 2018, 7(5):26-26. https://doi.org/10.1109/MCE.2018.2835880