

Design of music training assistant system based on artificial intelligence

Hua Zhihan¹, Liang Yuan², Tao Jin^{3,*}

¹Department of Art and Design, Shijiazhuang University of Applied Technology, Shijiazhuang, 050081, China;

²Department of Architectural Engineering, Shijiazhuang University of Applied Technology, Shijiazhuang, 050081, China

³Department of mathematics and Computer Engineering, Ordos Institute of Technology, Ordos, 017010, China

Abstract

In order to improve the input accuracy and response speed of music training, this paper designs an intelligent assistant system. The architecture is divided into infrastructure layer, data layer, application layer and presentation layer. In the hardware design, the combination of ARM and digital signal processor (DSP) is used to realize the interaction between data analysis and human and interface. In the software design, cepstrum algorithm is used to extract cepstrum features of music signals, linear smoothing algorithm is used to filter, dynamic time warping method is used to match patterns, and radial basis function algorithm is used to output the results. Thus, the overall design of the music-assisted training system is completed. Experimental results show that the signal-to-noise ratio of music signal transmission is more than 14dB, the accuracy is higher than 99.5%, and the response speed of serving 240 users is only 1s. The system has strong operability and good performance of music assistant training.

Keywords: Artificial intelligence; Music training; Assistance system; Audio acquisition module; Music signal; Radial basis function

Received on 22 December 2021, accepted on 05 February 2022, published on 11 February 2022

Copyright © 2022 Hua Zhihan *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.11-2-2022.173450

Corresponding author. Email: jintao@oit.edu.cn

1. Introduction

With the development of computer technology, the combination of computers and music has become an important trend. The music played by electronic instruments, electronic guitar, electronic organ and other instruments can be synthesized by synthesizer, and the favourite music can be compiled by computer [1-2]. It is imperative to build a music training assistant system based on modern information technology. The music training assistant system can not only realize the effectiveness of the music teaching system, but also realize online music knowledge transfer, student learning status investigation, assessment and performance analysis, as well as organic interaction between teachers and students, so as to provide an interactive system for teachers and students, promote students to learn music course knowledge easily and happily [3-4], and teachers can also fully understand students' actual needs.

In music teaching, designing a music training assistant system can improve students' sensory perception of music and stimulate students' interest in learning [5-6]. Computer technology has been widely popularized, which has laid a solid foundation for the application of music digital teaching assistant system in teaching practice.

Music training assistant system can let students experience a new education method and mode. Students' learning is no longer limited by time and space, and learning activities are mainly done by students themselves. The music training assistant system spreads the music training information through internet teaching by means of electronic documents, video and audio files and other carriers [7]. Using modern educational technology for teaching can expand the scope of education, break through the constraints of traditional space and time, and alleviate the current situation of the lack of software and hardware in running traditional colleges and universities to a great extent. The music training assistant system has the freedom, flexibility and openness of learning time, so that more students can learn independently according to their own needs, and can

provide more people with opportunities for learning and further study [8]. Teachers can upload and provide a large number of learning files and materials related to music, including corresponding folk music, opera music, folk songs, Quyi music, instrumental music, some folk sacrificial ceremony music and other teaching materials. For a large number of video and audio files, students can download and learn according to their own needs. For students with different learning abilities, they can carry out song jump learning according to their own needs, so that students can really learn independently and cultivate their autonomous learning ability [9].

At present, there is research into the combination of artificial intelligence technology and music. Tong applied artificial intelligence technology to the music sight singing guidance system [10], a computer system that uses computer to imitate the experience and methods of teaching experts to assist teaching, and provides learners with personalized learning resources and adaptive teaching methods through ITS. The research contents include the definition of music scoring difficulty characteristics, the design of music scoring recommendation algorithm, and the design of solfeggio scoring algorithm, and realized the effective guidance of music sight singing through artificial intelligence technology; Guo et al. applied artificial intelligence technology to language training strategy optimization system [11], using computer technology to build a teaching platform, embedding voice recognition technology into it, and using artificial intelligence technology to input, analyze and recognize voice, when it is applied to music teaching class, music can be effectively recognized, and the wrong segments are prompted; Takama et al. applied implicit feedback technology to music recommendation system [12], in order to treat the context information of music tracks as auxiliary information to compensate for implicit feedback, this paper adopts a decomposition machine with context information as its feature. Taking advantage of FMs's easy introduction of features, besides context features, this paper also introduces content features, carries out multi-plane reconstruction, and realizes clear information extraction of music, and the two methods effectively realized the good combination of artificial intelligence technology and music. At present, some researchers apply artificial intelligence technology to training. Ding applied artificial intelligence technology to public management teaching practice training performance analysis [13], introduce evaluation parameters and adopt analogy method to introduce artificial intelligence into the model proposed in this study. Combined with the actual demand of music-assisted teaching, a teaching effect analysis model based on artificial intelligence technology is constructed. B/S structure is used to construct the system, functional modules are set up based on demand analysis, and an expert system is used to evaluate the music teaching effect, so as to realize the good combination of artificial intelligence technology and teaching. However, the effect of applying artificial intelligence technology to music training assistant system is not ideal, and the accuracy of system input and transmission of music data is not high. In order to solve this problem and further meet the needs of music

intelligent teaching, this paper designs a music training assistant system based on artificial intelligence. The presentation layer of the system is responsible for managing the user operation interface and business operation interface; The data layer is responsible for the informatization of basic information and statistical data of song training; The application layer is responsible for managing training resources and sending user requests to the infrastructure layer; The main control module of the infrastructure layer processes the music signal through cepstrum algorithm and linear smoothing algorithm, uses the dynamic time warping method to match the speech signal mode, and uses the radial basis function algorithm to output the system training results. The system selects TMS320VC5502 as the DSP chip of the main control module and TLV320AIC23BPW as the audio processing chip of the system's audio acquisition module. Through the system test, it is verified that the designed system has high assistant performance of music training and can provide effective assistance for music teaching.

2. Materials and methods

2.1 Overall design of the system

The main function of the music-assisted training system based on artificial intelligence is to enable multi-level users to use the designed system to realize online music-assisted training, which makes music-assisted training more efficient, safe and has good human-computer interaction. The overall structure of music training assistant system based on artificial intelligence is shown in Figure 1.

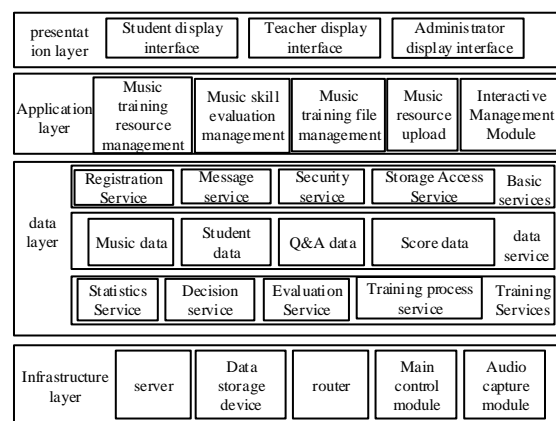


Figure 1. The overall structure of the system

The music training assistant system based on artificial intelligence mainly includes four parts: infrastructure layer, data layer, application layer and presentation layer.

Infrastructure layer

The infrastructure layer of the system includes various hardware devices set by the system, and the main control module is responsible for the processing and output of music signals and audio acquisition. This layer is the main support for cross platform application of the system and ensures the technicality and scalability of system software operation.

Data layer

The music training assistant system uses the data layer to realize the informatization of music training assistant statistics and management information. The data layer mainly realizes the data interaction and access of assistant management information of digital song training [14]. The data layer of the system will provide certain interfaces to facilitate system expansion or standby for other systems.

Application layer

The music training assistant system uses the application layer to realize music training resource management, music skill evaluation management, music training file management, music resource upload and interactive management.

Presentation layer

The presentation layer is the system interface presented to users. The presentation layer of the designed music training assistant system is mainly divided into client level and server level. The server level mainly assembles or encapsulates the functional modules developed by the system, and then presents them to users through the client. The client layer is mainly used and operated by users. Therefore, the server layer is equivalent to a container, while the client layer is mainly used by users [15]. The presentation layer is embodied in the management operation interface and business operation interface, as well as the logical relationship between interface operations.

After logging into the system through the display layer, the user selects various required music assisted training applications through the application layer. The service layer of the system provides users with various music assisted training services according to the music assisted training applications selected by the user, and the system uses the infrastructure layer to provide technical support for the system.

2.2 Hardware design

In order to meet the real-time requirements and improve the operation accuracy of the system, the system is composed of ARM and digital signal processor (DSP). The advantage of this scheme is that it combines the man-machine interface ability of ARM and the powerful operation function of DSP. The DSP in the system is mainly responsible for operation, and ARM is mainly responsible for signal processing and instruction issuing control.

Pitch and note duration are the two main elements of the main melody of music signal. In this system, the audio

signal acquisition module collects the music signal, and the main controller ARM receives the key signal and controls the DSP's selection function. If the automatic spectrum recording function is selected, DSP will calculate and process the collected music signal to identify the pitch and duration of notes, convert the music signal into MIDI format data and store it in the storage space of ARM. ARM can play back the collected audio signal by controlling MIDI chip, convert MIDI format data into music score (staff score) and drive LCD display; If the instrument tuning function [16] is selected, the DSP only detects the pitch of the music signal and does not execute the note time value detection algorithm. The obtained pitch result is sent to ARM and displayed to the user by LCD, and the user will feed back and control the pitch adjustment direction of the instrument according to the displayed pitch value.

The system design comprehensively considers the factors such as volume, power consumption and real-time, and selects the chip. The core algorithm processor is TMS320VC5502, the main controller of the system is ARM S3C44BOX, the LCD is LM057QCIT01 of SHARP Company, the audio ad is 24 bit special voice analogy-to-digital conversion chip CS53L21 with low-power, and the MIDI chip is MC-KED-001 with excellent synthesis ability. The hardware structure of the system is shown in Figure 2.

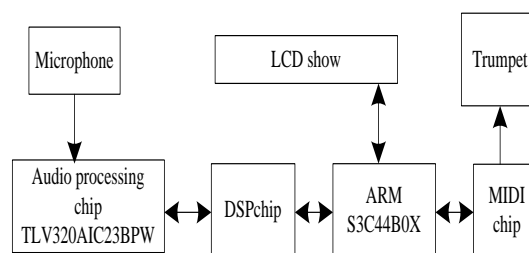


Figure 2. Diagram of system hardware structure

As can be seen from Figure 2, the system adopts ARM and DSP to form the hardware system. The ARM main controller is responsible for displaying the system interface, controlling the communication connection of multiple core chips such as DSP, and managing the transmission clock of serial data; The DSP control chip is responsible for the signal processing part, and the audio codec TLV320AIC23BPW is responsible for the audio acquisition part.

DSP control chip

The DSP control chip is located in the main control module of the system infrastructure layer. The core operation processor of the system is TMS320VC5502. DSP chip, also known as digital signal processor, is a microprocessor with special structure. The internal structure of DSP chip adopts Harvard structure with separate program and data, has special hardware multiplier, widely adopts pipeline operation, provides special DSP instructions, and quickly

realizes various digital signal processing algorithms.

TI's DSP chip has the characteristics of high performance, good stability and a wide variety, mainly including fixed-point and floating-point. The advanced power management technology of TMS320C55x series DSP can automatically turn off idle peripherals, memory and core functional units, extending the battery life [17]. For the music training assistant system, low power consumption and real-time performance are the main performance requirements. TMS320VC5502 is a high-performance, low-power fixed-point digital signal processor. The main performance indexes are as follows:

The main features are:

(1) 3.34-6ns instruction cycle, 30-200MHz clock frequency, 16K byte instruction cache (I-Cache), 1-2 instructions can be executed in each cycle, with double multiplier, 2 arithmetic logic units, 1 program bus, 3 internal data / operand read buses and 2 internal data / operand write buses.

(2) 32K word on-chip RAM (composed of 8 4K word dual access RAM): 16K word on-chip ROM in waiting state; Maximum addressable external memory space of 8m words.

(3) External memory interface (EMIF) for general input and output supported by 32-bit parallel bus memory: interface that can be connected with asynchronous SRAM, asynchronous EPOM, synchronous DRAM, synchronous Burst RAM, etc.

(4) Hardware simulation and debugging a tracking ability, which can save the first 16 program counter interrupts and the first 32 PC values; Programmable low-power control can be carried out for the functional domain of six devices.

(5) On-chip hardware-based logic scanning; 1149.1 boundary scan logic with IEEE standard: 176 pin LFPF and 201 pin BGA; 3.3 VI/O supply voltage and 1.35V core voltage.

TMS320VC5502 chip includes 6 channels of memory direct access controller; 3 multi-channel buffered serial ports; Programmable analogy PLL clock generator; General purpose I / O pin (GPIO) and special output pin (XF); 8-bit / 16-bit parallel host interface; 4 timers: two 64-bit general purpose timers, one 64-bit programmable watchdog timer, one 64-bit DSP/BIOS counter; 12C interface; Universal asynchronous transceiver (UART); TMS320VC5502 has 16MB storage space. When a certain space is used as a data space [18], the minimum addressing unit is 16-bit; When a certain space is used as program space, the minimum addressing unit is 8-bit.

Audio acquisition module

The audio acquisition module of the system is located in the infrastructure layer of the system. The audio processing chip of the system adopts the audio codec TLV320AIC23BPW of TI company. It has built-in headphone output amplifier. Both input and output have programmable gain adjustment function and low power consumption. This makes it an ideal audio processing chip. TLV320AIC23 is connected to the control MCU of the system through McASP. TLV320AIC23 is connected with DSP through control interface and digital

audio interface. The control interface is used to set the working parameters of AIC23; Digital audio interface is used for audio data transmission between TLV320AIC23 and DSP chip. The connection structure between TLV320AIC23 and DSP chip is shown in Figure 3.

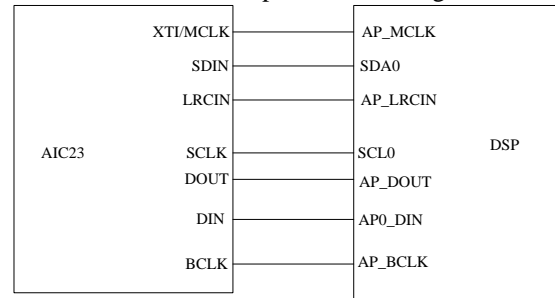


Figure 3. TLV320AIC23 and DSP chip connection structure diagram

The BCLK in Figure 3 is the serial data transmission clock. When the TLV320AIC23BPW is in the master mode, the BCLK is generated by itself and provided to the DSP. The frequency is 1 / 4 of the master clock. When the AIC23 is in the slave mode, it is generated by the DSP; DIN is the serial data input terminal, which is sent to the stereo DAC, DOUT is the serial data output terminal, which is generated by the stereo ADC; XTi / MCLK is the external clock input, which is used to generate the internal clock of AIC23.

2.3 Software design

There are high-frequency and low-frequency signals in music. In order to accurately identify music signals, a cepstrum method is proposed to extract the frequency features of music signals. The features obtained at this time are called cepstrum features. Pitch period is one of the most important parameters in speech signal. In order to deeply identify speech signal, it is necessary to analyze the spectrum information of pitch in the obtained cepstrum features.

The software design of the system mainly uses cepstrum algorithm to extract the cepstrum characteristics of music signals, and uses linear smoothing algorithm to filter the music signals; The dynamic time warping method is used to match the speech signal pattern, and the radial basis function algorithm is used to output the system training results.

Cepstrum feature extraction based on cepstrum method

Among the pitch detection algorithms, cepstrum method is a common algorithm. Its principle is to detect some pitch information based on the cepstrum features of speech signal, which mainly represents the glottic excitation period. With the support of channel response filtering $v(n)$ and glottic

pulse excitation $e(n)$, speech $s(n)$ can be obtained [19], that is:

$$s(n) = e(n) * v(n) \quad (1)$$

It can be seen from the above formula that the short-time spectrum of sound signal can be characterized by the excitation source spectrum and filter spectrum, that is, the product of the two. In the short-time spectrum of voiced signal, the periodic fine structure changes rapidly, and its corresponding things are the fundamental frequency and harmonics of periodic pulse excitation. In the frequency domain, the information of the source and the channel is still mixed together and does not exist independently. Let $\hat{s}(n)$, $\hat{e}(n)$ and $\hat{v}(n)$ be the cepstrum of $s(n)$, $e(n)$ and $v(n)$ respectively, which can be obtained from the properties of convolution and Fourier transform:

$$\hat{s}(n) = \hat{e}(n) * \hat{v}(n) \quad (2)$$

It can be seen from the above formula that in cepstrum domain, pitch signal and channel information exist independently and can be separated. With the support of simple cepstrum filtering method, $e(n)$ and $v(n)$ can be segmented. In this case, the pitch period can be obtained from the separated excitation $e(n)$ and cepstrum features.

Linear smoothing algorithm

The linear smoothing process uses the sliding window for linear filtering, that is:

$$y(n) = \sum_{m=-L}^L x(n-m) * w(m) \quad (3)$$

Where $\{w(m), m = -L, -L+1, \dots, 0, 1, 2, \dots, L\}$ is

the smooth window of point $(2L+1)$, which satisfies:

$$\sum_{m=-L}^L w(m) = 1$$

. Linear smoothing can not only minimize the value of "wild points" in the signal, but also modify the value of samples to a certain extent. In this case, when the length of the added window increases, the smoothing effect is more obvious, that will also aggravate the step phenomenon between the two adjacent smoothing segments and make the signal more blurred. In order to make the smoothing effect more obvious, the two median values are connected, and the combination of the two median values is smoother.

Speech signal pattern matching based on dynamic time warping

After preprocessing and extracting speech feature parameters, pattern matching is needed to realize speech recognition. Dynamic time warping (DTW) is a nonlinear

warping method which combines time warping with distance gap calculation. Through this algorithm, the problem of unequal time length of speech signal feature parameter sequence comparison can be successfully solved, which makes it easier for speech features to match with standard values, thus improving the accuracy of recognition. Let the feature vector sequence of the standard reference template be

$$R = \{R(1), R(2), \dots, R(m), \dots, R(M)\},$$

where M is the total number of speech frames contained in the standard reference template, m is the timing label of the

frame in the standard reference template, and $R(m)$ is the feature vector of the speech in frame m . The feature vector sequence of the input speech test template is $T = \{T(1), T(2), \dots, T(n), \dots, T(N)\}$, where N is the number of frames of speech in the test template, n is the

timing label of speech in the template, and $T(n)$ is the feature vector of speech in frame n . The matching similarity between the feature vectors of the test template and the reference template can be expressed by the distance between the vectors. The greater the distance is, the smaller the matching similarity is. The distance between feature vectors $T(n)$ and $R(m)$ is usually expressed by Euclidean distance:

$$d[T(n), R(m)] = \sum_{i=1}^p (t_i - r_i)^2 \quad (4)$$

Where t_i and r_i represent the i -th dimensional eigenvectors of $T(n)$ and $R(m)$ respectively, and P is the order of the eigenvector. The dynamic time warping algorithm is to find the time warping function $m = w(n)$, and nonlinearly map the time axis n of the test template to the time axis m of the reference template, so as to minimize the total matching distance D between the test template and the reference template, that is:

$$d = \min_{w(n)} \sum_{n=1}^N d[T(n), R(w(n))] \quad (5)$$

DTW algorithm is an optimization problem. It uses the time warping function $w(n)$ to describe the time correspondence between the test template and the reference template, so as to solve the matching distance between the two templates in this case. Usually calculated by dynamic programming method (DP), the dynamic time warping algorithm is to find a path from the starting point to the ending point and passing through each intersection, so as to minimize the sum of frame distance measures of all intersections on the path [20]. The speed of language pronunciation will vary, but the pronunciation order of each phoneme will not change. Therefore, this path must start at

the lower left corner and eventually lead to the upper right corner. Constraint conditions are given for the slope of the path at each point:

Boundary conditions:

$$w(1) = 1, w(N) = M \quad (6)$$

Continuous conditions:

$$w(n+1) - w(n) = \begin{cases} 0, 1, 2 & w(n) \neq w(n-1) \\ 1, 2 & w(n) = w(n-1) \end{cases} \quad (7)$$

The dynamic time planning problem boils down to finding the optimal path function $m = w(n)$ on the premise of meeting the constraints, so that the sum of the accumulated frame distances along the path is the minimum.

Starting from point $(1,1)$, for a point on the path, the previous lattice point reachable by point (i_n, i_m) can only be point $(i_n - 1, i_m)$, point $(i_n - 1, i_m - 1)$ and point $(i_n - 1, i_m - 2)$. Then point (i_n, i_m) must select the point with the smallest vector accumulation distance among the three points as the preceding grid point. The path vector distance accumulation is:

$$D(i_n, i_m) = d(T(i_n), R(i_m)) + \min\{D(i_{n-1}, i_m), D(i_{n-1}, i_{m-1}), D(i_{n-1}, i_{m-2})\} \quad (8)$$

Let $D(1,1) = 0$. Starting from point $(1,1)$, the search is repeated until (N, M) , and the best matching path can

be obtained. $D(N, M)$ is the total matching distance of the two template vectors corresponding to the best matching path. When the system performs speech recognition, the minimum matching distance $D_{\min}(N, M)$ is obtained by matching the test template and the reference template. This minimum matching distance can be used as a measure of pronunciation similarity between the reference template and the test template, which can reliably and comprehensively reflect the similarity of language features.

Music training results

In order to avoid the error in matching, the radial basis function neural network is used to compensate the error of identification model in time. The radial basis function algorithm in artificial intelligence technology is selected to output the system training results. Radial basis function algorithm is a neural network composed of locally adjusted neurons, which is a five layer network model. It belongs to the type of forward neural network, which can approximate any continuous function with any accuracy, and is especially suitable for solving classification problems. The first layer is the information factors related to the case. These inputs can be summarized into different music project indicators and input into the neural network structure. The second layer is

the membership function, the input item is the music item index, and the membership degree of the calculated output item index, and its mathematical expression is as follows:

$$\mu_{ij}(x_i) = \exp\left[-\frac{(x_i - c_{ij})^2}{\sigma_j^2}\right] \quad (9)$$

where, $i = 1, 2, \dots, r$, $j = 1, 2, \dots, u$.

The third layer defines the number of fuzzy rules, and makes the number of learning rules the least and most important through sample learning. The output of rule j is mathematically calculated as follows:

$$\phi_j = \exp\left[-\frac{\sum_{i=1}^r (x_i - c_{ij})^2}{\sigma_j^2}\right] = \exp\left[-\frac{\|X - C_j\|^2}{\sigma_j^2}\right] \quad (10)$$

Where, $c_j = (c_{1j}, \dots, c_{rj})$ represents the center of the j -th RBF unit. The feature of RBF neural network is that the closer the neuron is to the center, the higher the activation degree, and the deeper the features of music signals can be.

The fourth layer is the normalization layer. The nodes of this layer are consistent with the fuzzy rule nodes. The output formula of the j -th node N_j is as follows:

$$\psi_j = \phi_j / \sum_{k=1}^N \phi_k \quad (11)$$

The fifth layer is the output layer, which outputs the evaluation results of the music training assistant system. The output is as follows:

$$y(x) = \frac{\sum_{i=1}^n \left[(a_{i0} + a_{i1}x_1 + \dots + a_{ir}x_r) \exp\left(-\frac{\|x - c_i\|^2}{\sigma_i^2}\right) \right]}{\sum_{i=1}^n \exp\left(-\frac{\|x - c_i\|^2}{\sigma_i^2}\right)} \quad (12)$$

w_k represents the connection mode of the k -th rule, that is, the sum of the weight product of the output variable is expressed as follows:

$$y(x) = \sum_{i=1}^n w_k \cdot \psi_k \quad (13)$$

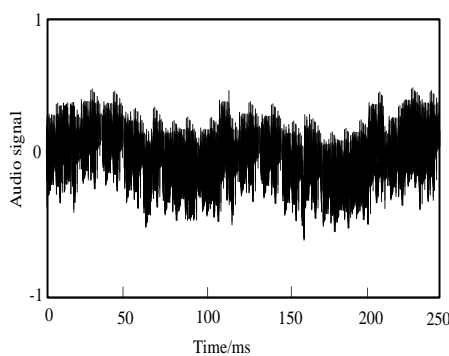
In the above calculation process, X is the proportion of time consumed by students with excellent music training results in the music training assistant system; $c_j \sigma_j$ is the distribution of hidden layers of each music learning courseware for the students with excellent results in the corresponding X learning time. The hidden layer categories of these learning samples are relatively parallel; Y is the best music training score corresponding to each input layer. This five-layer network model divides the input music items through the first layer, the second layer is to calculate the membership degree of the music items and determine the matching of the input music signals in a fuzzy

way, the third layer is to determine the number of fuzzy judgment rules, and the fourth layer is the node for matching the input music signals and rules, among which the second layer, the third layer and the fourth layer are all to determine the matching of the input music signals in a fuzzy way, and the fifth layer is to output the matching results and complete the training evaluation. In order to simplify the RBF algorithm, the second to fourth layers are reduced to hidden layers. The first and fifth layers are input layer and output layer respectively. According to different training directions of music teaching, they are used as an algorithm's input layer to output training evaluation results.

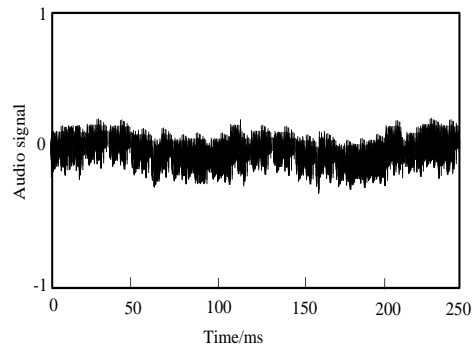
3. Results

After the design of music training assistant system based on artificial intelligence is completed, it needs to meet the accuracy requirements of data query and the operability of concurrency control. The system needs to have high response performance and can meet many needs such as query, request, calculation and so on. In order to test the assistant performance of the music training assistant system for music training, the designed system is applied to a music college, and the feasibility of the designed system is tested through the system operation. Windows Core i7 processor, 16GB memory, 2.20Hz flash memory and 1TB hard disk are used as the hardware configuration of the experiment. Use the Last.fm music data set, randomly select 500 pieces of music to train the system, set the maximum iteration number to 5000 times, the initial learning rate to 0.1, and the learning rate change factor to 0.2. After the training is completed, select 200 pieces of music from the remaining music to form a test set for testing.

In this paper, the signal filtering before and after filtering of music signal by linear smoothing algorithm is counted, and the statistical results are shown in Figure 4.



(a)Before filtering



(b)After filtering

Figure 4. Music signal changes before and after filtering

As can be seen from the comparison results before and after filtering in Figure 4, the system in this paper can effectively filter the music signal and realize the effective smoothing processing of the music signal in the system. The linear smoothing algorithm used in this system uses the sliding window to linearly filter the music signal, which has high filtering effectiveness. The fluctuation amplitude of the filtered music signal is about 43% smaller than that before filtering. The filtered music signal effectively avoids the interference of noise, improves the transmission quality of music signal in the system, and provides a signal basis for the good training of music training assistant system.

The signal-to-noise ratio of transmitted music signal in the designed system is counted. In order to intuitively show the filtering performance of this system, this system is compared with the system in reference [10] and the system in reference [11]. The comparison results are shown in Figure 5.

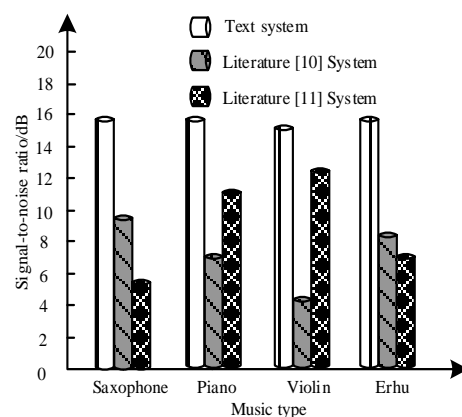


Figure 5. Comparison of signal-to-noise ratio of music signals in different systems

As can be seen from the system comparison results in Figure 5, The average signal-to-noise ratio of the music

signal output by the system in this paper is 15.2dB, while the average signal-to-noise ratios of the music signals output by the system in reference [10] and the system in reference [11] are 7.6dB and 8.2dB, respectively. It can be found that the system in this paper can effectively improve the quality of music signals. make the music signal transmitted by the system have high definition, and provide the basis for the accurate assistant training of the system for different types of music.

In this paper, the system in this paper uses the DTW algorithm to achieve good matching of music signals in the system, and sets the threshold of system matching signal to 3. When the user uses this system for audition, when the DTW result is higher than 3, it indicates that the system's user audition is successful. When the user uses the system of this paper, the DTW results obtained are shown in Figure 6.

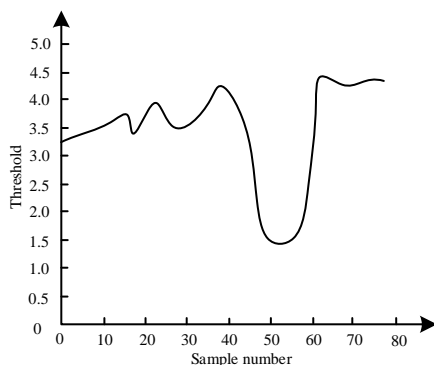


Figure 6. DTW calculation results

As can be seen from the experimental results in Figure 6, the matching threshold of the system in this paper is higher than 3, which indicates that the system's user audition is successful, and the DTW algorithm can be effectively used to realize the effective upload of music sight singing in the process of user training. The DTW algorithm used in this system has good matching performance, which can match the music uploaded by users with the standard music, and improve the application performance of the system.

The input function of the music assisted training system is very important. The music input function of the designed system is tested. This system is used to input the music played by piano, erhu, saxophone and violin, and test whether the designed system can correctly identify and input the corresponding music. 10 pieces of music played by different types of musical instruments are selected to count the note entry accuracy of music played by different types of musical instruments, and compare the designed system with the systems in reference [10] and reference [11]. The comparison results of music entry accuracy are shown in Figure 7.

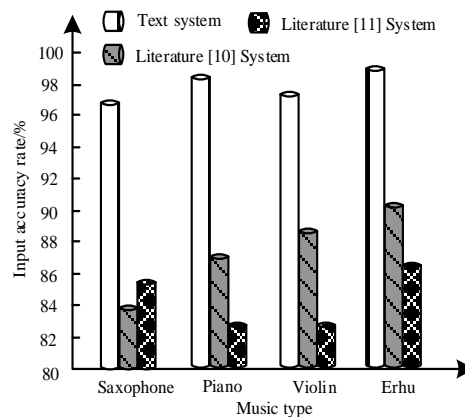


Figure 7. The correct rate of music recording

From the system test results in Figure 7, it can be seen that the system can effectively input music played by different musical instruments, and the input accuracy of music played by different musical instruments is higher than 97.5%, While that accuracy rate of music input in literature [10] system and literature [11] system are 86.2% and 84.1% respectively, indicating that the system in this paper has very high music input accuracy. This is because this method adopts efficient signal processing technology, uses the data processing performance of artificial intelligence technology, effectively realizes the assistant training in the process of music teaching and the good combination of music and computer, and has high application performance.

Statistics of the test results for various music assisted training operations of the designed system during the operation of the system are shown in Table 1.

As can be seen from the system test results in Table 1, the system in this paper can successfully complete various functions of music assisted training such as uploading music, score display and training course management. The operation results can meet the use needs of users. The test results of various test contents are successful, which verifies that the system has high operability and feasibility. The designed system has high effectiveness and can be applied to music assisted training.

The management and interaction between the client and background database of the system should have good concurrency. The completed system should be able to well support at least 150 clients to execute and operate the response process of a request at the same time, so as to meet the basic requirements of customers for batch data processing of the system. In terms of system processing and response, the background can quickly respond to the operations carried out by the client user interface, including query, request, calculation and other basic operations. The response speed of each operation of the designed system is counted when the concurrent users are 240. The statistical results are shown in Figure 8.

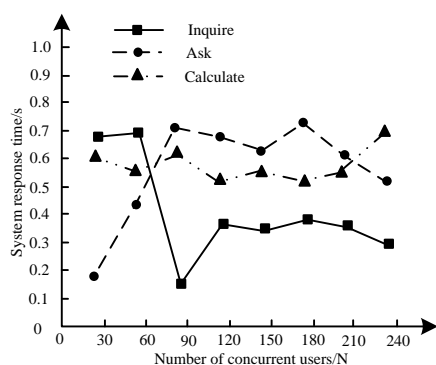


Figure 8. System response time for different concurrent users

Table 1. System running test results

Test serial number	Test content	User action	Operation result	Test Results
1	Reset operation	Reset after entering a random number	The system can quickly restore the original state	success
2	Login operation	Enter username and password	Successfully logged in to the system	success
3	Elective	Select the required music course	Successfully select the required course	success
4	Change information	Enter the changed information	Complete the change	success
5	Course test	Submit after completing the answer	Test successfully and output the result	success
6	Training course management	User Management Training Course	Complete management operations	success
7	Upload music	Click upload	Upload music completed	success
8	Question query	Enter key words for test questions	Questions pop up successfully	success
9	Results display	Choose the grade you want to display	Pop-up results	success
10	Recommended features	Click "Music Knowledge"	Show recommended results	success

As can be seen from the experimental results in Figure 8, when the designed system has 240 concurrent users, the calculation speed, response speed and response speed of the system are less than 1s, which verifies that the system has high data processing performance, can quickly process user requests and has good concurrent operability. The designed system can be very friendly to support various user operation requests, and quickly feedback the processed results to the display interface of the system in time, which is convenient for users to consult. With the continuous expansion and accumulation of data in the system, the access volume of the system and the calculation and operation volume of the front desk are increasing. It should be ensured that the system can respond within 30 seconds after the user's operation is issued. In terms of system

backup and recovery operation, the completed system should have automatic regular and irregular backup function and rapid recovery ability. Once the system encounters faults or corresponding problems, the system can make corresponding response within 24 hours, recover the system in time, and ensure the smooth operation of the system. The designed system can meet the basic needs of music training. It should be continuously improved and updated in the future training practice, in order to provide maximum help for music training.

4. Discussion

The experimental results show that the music-assisted training system based on artificial intelligence technology

designed in this paper, the signal-to-noise ratio of the output music signal reaches 15.2dB, and the average accuracy of music input is 97.5%, which is far superior to other comparison systems. This shows that this system can improve the effect of music-assisted training and provide the basis for students to learn music and teachers to teach music. The designed music training assistant system includes four functional layers: infrastructure layer, data layer, application layer and presentation layer. Through man-machine design, it ensures that users can operate directly on the client and input the relevant address of the system on the home page of the browser to meet the needs of system operation. When users use the system for data input, the foreground and background of the system judge and detect the input information submitted by users, record the correct information format into the system database, reduce the occupation of database capacity by invalid information, reduce database memory requirements, and improve the operation speed and efficiency of the system. All data exchange in the system is carried out in the network, which reduces the task and pressure of the client, user's manual input and user's operation input errors, and ensure the accuracy of data. Because the system handles and captures many exception handling situations, the later system development and maintenance are relatively simple. The system does not need complex business processes, and subsequent maintenance and operation personnel only need to carry out daily management and monitoring of the system. Especially in today's era of popularization of computer technology, they can adapt to system management after short-term training. From the perspective of system operation, it has considerable practical value.

5. Conclusion

In order to improve students' music training effect and enhance students' cognitive level of music, modern technology is introduced and artificial intelligence is applied to build a music-assisted training system. In the system, through hardware and software design, DSP and ARM processor are used to realize signal analysis and command control of each component in the music system, cepstrum analysis method is introduced to dig deeply into the music signal, identify the characteristics of the music signal, and five-layer network model is used to realize the evaluation and output of music, thus completing the design of music-assisted training system. The test results show that the system runs well, can meet the existing design goals to a great extent, can meet the needs of students and teachers for music-assisted training, and improves the efficiency and quality of music teaching to a certain extent. However, the interaction of the system is not considered enough in the application of this system. In the future research, it is necessary to consider adding related content such as interface design to provide more powerful support for practical application..

Acknowledgments

This study is supported by scientific research project of colleges and universities in Inner Mongolia Autonomous Region with No.NJZY21153, Ordos Science and technology planning project with No.2021YYI18-46, Higher Education Reform Project of Ordos Applied College with No.KYYB2019007.

References

- [1] [ESLING P, MASUDA N, BARDET A, et al. Universal audio synthesizer control with normalizing flows[J]. Applied Sciences, 2019, 10(1):302.
- [2] PITT I L. Life cycle effects of technology on revenue in the music recording industry 1973–2017[J]. SN Business & Economics, 2020, 1(1): 9-10.
- [3] ANUFRIEVA N I, KAMENETS A V, PEREVERZEVA M V, et al. Features of the application of art-therapeutic and gaming technology based on folk music in rehabilitation and socialization of children with health limitations[J]. Open Access Macedonian Journal of Medical Sciences, 2020, 8(E): 373-381.
- [4] REBROVA O. Artistic and aesthetic innovations of art education used while training future teachers of music and choreography[J]. Scientific bulletin of South Ukrainian National Pedagogical University named after K D Ushynsky, 2019, 4(129):13-25.
- [5] DUIC L. Choose music! a consulting and training strategy for admission to higher education in music[J]. Review of Artistic Education, 2019, 17(1):40-46.
- [6] HU W J. Algorithm for sampling outliers in imbalanced data sets of artificial intelligence[J]. Computer Simulation, 2019, 37(11):324-328.
- [7] WOLFFENBÜTTEL C R. Music education training for teachers[J]. Creative Education, 2019, 10(10):2101-2110.
- [8] LIU S, WANG S, LUI X, et al. Fuzzy Detection aided Real-time and Robust Visual Tracking under Complex Environments[J]. IEEE Transactions on Fuzzy Systems, 2021, 29(1):90-102.
- [9] PEREVERZEVA M V, OREKHOVA O G. Music hermeneutics, semantics, semiotics and their role in professional training of the musician[J]. Uchenye Zapiski RGSU, 2019, 18(4):124-130.
- [10] LIU S, LIU X, WANG S, et al. Fuzzy-Aided Solution for Out-of-View Challenge in Visual Tracking under IoT Assisted Complex Environment[J]. Neural Computing & Applications, 2021, 33(4): 1055-1065
- [11] GUO J, LIU J. Optimal system design of language training strategy based on artificial intelligence[J]. Journal of Intelligent and Fuzzy Systems, 2020, 40(4):1-11.
- [12] TAKAMA Y, ZHANG J C, SHIBATA H. Context-aware music recommender system based on implicit feedback[J]. Transactions of the Japanese Society for Artificial Intelligence, 2021, 36(1):1-10.
- [13] DING Y. Performance analysis of public management teaching practice training based on artificial intelligence technology[J]. Journal of Intelligent and Fuzzy Systems, 2020, 40(5):1-14.
- [14] KOEMPEL F. From the gut? questions on artificial intelligence and music[J]. Queen Mary Journal of Intellectual Property, 2020, 10(4):503-513.

- [15] AVDEEFF M. Artificial intelligence & popular music: skygge, flow machines, and the audio uncanny valley[J]. Arts, 2019, 8(4):130.
- [16] WILLIAMS R. Artificial intelligence assistants in the library siri, alexa, and bevond[J]. Online, 2019, 43(3):10-15.
- [17] STERNE J, RAZLOGOVA E. Machine learning in context, or learning from landr: artificial intelligence and the platformization of music mastering[J]. Social Media + Society, 2019, 5(2):205630511984752.
- [18] JABADE V, DESHPANDE V, ADITYA K. Music generation and song popularity prediction using artificial intelligence - an overview[J]. International Journal of Computer Applications, 2019, 182(50):33-39.
- [19] SHUAI L, TENGHUI H, JIANHUA D. A Survey of CRF Algorithm Based Knowledge Extraction of Elementary Mathematics in Chinese[J]. Mobile Networks & Applications, 2021, 26(5):1891-1903
- [20] PENG G, JINGYI :L, SHUAI L. An Introduction to Key Technology in Artificial Intelligence and big Data Driven e-Learning and e-Education[J]. Mobile Networks & Applications, 2021, 26(5): 2123-2126