

## Research Article

# Smart Speech Recognition System for Chinese Language Learning Enhancement

Lijun Hu<sup>1</sup> and Jing Jia<sup>2</sup> 

<sup>1</sup>*Puyang Medical College Puyang, Puyang, Henan 457000, China*

<sup>2</sup>*Faculty of Linguistic and Cultural Studies, Changchun Sci-Tech University, Changchun, Jilin 130000, China*

Correspondence should be addressed to Jing Jia; 100507@cstu.edu.cn

Received 25 April 2022; Revised 16 May 2022; Accepted 20 May 2022; Published 10 June 2022

Academic Editor: Lianhui Li

Copyright © 2022 Lijun Hu and Jing Jia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the expansion of teaching scale and the rapid development of educational information, a smart classroom management system based on speech recognition has been proposed and developed to improve the information level of smart classroom management. This paper discusses the application of multimedia equipment control. The system relies on the mature campus network in the way of cloud and local speech database for speech recognition. The application of the system proves that the smart classroom management system based on speech recognition in cloud architecture has more advantages than the traditional multimedia classroom management system and also has definite expansibility. It facilitates the unified management of the school, improves the efficiency of administrators, saves a lot of human and financial resources, and greatly promotes the development of school information construction. The smart classroom becomes a new direction in the application of information technology in the field of teaching and learning and is a smart learning scenario for improving the learning and teaching scenario. This paper focuses on the use of smart classroom as a supplementary teaching tool to improve Chinese language teaching and learning and illustrates the optimization of the teaching environment by smart classroom in terms of vocabulary learning and listening and speaking training.

## 1. Introduction

With the rapid development of modern computer science and technology, the traditional multimedia classrooms in schools have been continuously improved and upgraded to gradually form intelligent classrooms [1, 2]. The management of multimedia classrooms, from the traditional manual manipulation of various multimedia equipment to centralized control through the central control, then to remote control through the campus network, and then to the current automated management based on artificial intelligence, fully illustrates that the development process of education gradually began to apply artificial intelligence, schools focus on the construction of smart classrooms, and smart campus has become the future trend [3, 4]. Domestic construction of smart classroom design concepts and teaching models is relatively abundant but still lacks comprehensive practice [5].

The Ministry of Education proposes to promote the application of emerging technologies such as cloud computing and big data in school education and vigorously build education modernization. Regarding the active promotion of the application of “Internet+,” the guidance of the State Council marks the new technological revolution towards today’s stage [6, 7]. Therefore, it is necessary to develop a smart classroom management system based on speech recognition with cloud computing architecture. Smart campus is a comprehensive system implicating technologies from multiple fields such as cloud computing, campus network, big data, and remote control, and thus, it can only serve the students and teachers better after adequate integration and collaborative work [8, 9]. The basis for the realization of a smart campus is the Internet of Things (IoT), which relies on numerous application service systems to integrate teaching management, academic research, and

campus life of teachers and students and ultimately construct an integrated and intelligent campus environment for work, study, and life [10]. Smart classroom is the most important part of building smart campus, and it is the key work for universities to realize the strategic goal of informationization. At present, most schools have the foundation of campus network, especially some colleges and universities, and the campus network has been quite mature after many new constructions, upgrades, and renovations in the school informatization construction, which has laid a good foundation for the realization of smart classroom [11, 12]. Based on the modern technology, it is feasible to design an intelligent classroom management system by taking advantage of the existing campus network, cloud computing, and local customized voice library.

With the deepening application of big data, cloud computing, Internet of Things, and artificial intelligence technologies in education in recent years, the smart classroom has emerged as an overall solution to enhance teaching effectiveness [13, 14]. The optimization of the Chinese learning environment by smart classroom is reflected in three aspects: (i) accurate content pushing, (ii) efficient environment management and resource acquisition, and (iii) contextual setting and interactive feedback. Teachers can use the environment of smart classroom to effectively organize and manage the increasingly abundant teaching resources so as to better interact and teach students according to their aptitude.

Chinese is the foundation of language and has a significant impact on the development of habits and interest in future language learning. The use of smart classrooms to create interesting and shade-appropriate learning environments for students is the focus of this paper.

## 2. Technical Principle

The object of speech recognition research is speech, which is processed first, and the human voice is automatically recognized and understood by pattern recognition computers [15]. The combination and codevelopment of cloud computing and big data have contributed to the advancement of speech recognition technology to some extent. The deployment of deep learning framework in the cloud can enhance the capability of cloud computing, so the mutual promotion of deep learning, big data, and cloud computing greatly improves and enhances the ability of speech recognition models to mine and learn from complex data [16, 17]. Speech recognition systems are mainly divided into three types: embedded speech recognition systems, server-mode speech recognition systems, and cloud computing-mode speech recognition systems [18, 19]. According to the characteristics of the school, the human-computer interaction module of the intelligent classroom management system is based on speech recognition technology, while the piece of speech recognition based on cloud mode is not mature enough, and the customizable speech recognition service provided by the service provider is still at a primary level, so the speech recognition module of this system has to be divided into two parts: the speech cloud and the local

speech library. The voice cloud is responsible for the daily chat function, and the self-developed local voice library is responsible for the recognition of multimedia device control commands. The local speech library can be customized to provide speech recognition services for multimedia device control in the classroom, focusing on the reduced recognition range of these device controls and achieving higher recognition rates. The implementation of the local speech library requires the download and installation of the Microsoft Speech Recognition library, and the engine of speech recognition is driven by the speech recognition engine that comes with Windows, which can realize the ready acceptance of commands issued by the user [20].

This design currently drives three LCD displays simultaneously through this LCD driver all-in-one board; combined with the Android system software, it can realize the display of the conference theme or speaker's personal information on the upper bar screen, the lower standard screen displays the conference content, company logo, or video information, and the back bar screen carries the intelligent teleprompter system, which can display the speaker's speech content in real time.

*2.1. Hardware Circuit Design.* The main circuit board is a multimedia network player-LCD driver board based on Rockchip main chip RK3288. The design board can support LVDS/EDP/MIPI dot screen+HDMI dual display, LVDS+EDP dual display, LVDS+MIPI dual display, EDP+MIPI dual display, and other dual display mode options. It can drive 7-100-inch LCD and can support 4K full HD video decoding and 3840 × 2160 (for VOP\_BIG) and 2560 × 1600 (for VOP\_LIT) TFT LCD, and the main board contains 2 RS232, 2 UART, 4 USB HOST, 1 Ethernet, and other interfaces. It has a powerful communication function.

The design currently drives three LCD displays through this LCD driver board; combined with the Android system software, it can display conference theme or speaker personal information on the upper bar screen, the lower standard screen displays conference content, company logo, or video information, and the back bar screen equipped with intelligent teleprompter system can display the speaker's speech content in real time.

The smart podium consists of the following main parts: support structure, front face, operating table, LCD, and teleprompter. The usage frequency of each component in different application scenarios is shown in Figure 1.

The smart podium involves a total of three LCD displays, including two 28" bar screens and one 23.8" standard screen. The LCD module includes a display area, a PCB board set above the display area, and a COF (crystal coated film) set between the display area and the PCB board, the display area includes a protective film set at the front, a polarizer behind the protective film is defined as the reserved area, and the other defective area is defined as the defective area. The COF is fixed first, then the protective film and polarizer in the defective area are removed, the defective area in the display area is cut, then the lower end of the display area remaining after cutting is sealed, and finally, the

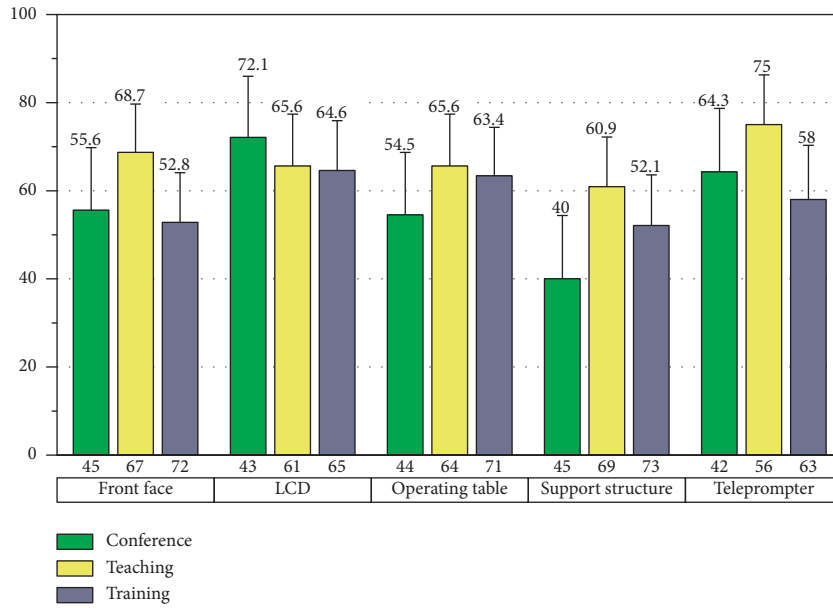


FIGURE 1: The usage frequency of each component in different application scenarios.

resolution of the sealed product is adjusted, so that the processed product can be used as a new small-size strip screen, thus achieving a novel display effect and being more widely used.

**2.2. Software Function Design.** The intelligent platform control software includes the information release of the first two display screens and the content display of the teleprompter system. The user can determine whether the device is online through the display status at the top of the software. Users can personalize the content of the two displays according to actual needs, including the display of pictures, videos, text font color size, and the interval of each material switch.

At the same time, the software also supports the preview function before the release of information to ensure that the information is accurately delivered to the audience, which greatly enhances the intelligence and personalization of the product, and can make the product applicable to a variety of different speech processes, which frequently can be calculated by the following equations. According to the formula (1)~(3), four teaching scenarios are selected for calculation, and the results are shown in Figure 2. It can be seen from the figure that, with the improvement of product intelligence and personalization, the audience will receive more information.

The intelligent teleprompter system software can automatically read the file content in the USB flash disk through the USB interface of the podium desktop for the speaker to deliver a speech. The operation of the speech content includes two modes: manual and automatic. In the manual mode, the speaker uses his or her voice to click the mouse to turn the page. The automatic mode is split into two categories: one is to set the automatic scrolling screen according to the speaker's personal preferences and reading speed for personalized settings, and the second is through the voice

recognition technology and teleprompter system linkage, so that the speech does not read and does not go, it has a read mark and is perfectly timed with the speaker's speech.

$$n = \sum_{i=1}^n n_i,$$

$$E_n = \sum_{m=0}^{N-1} x_n^2(m), \quad (1)$$

$$M_n = \sum_{m=0}^{N-1} |x_n(m)|,$$

where  $E_n$  is the amount of accurate information delivered to the audience;  $M_n$  is the degree of intelligence and personalization of products.

**2.3. Speech Recognition System Design.** The main function of the speech recognition program is to identify the voice commands that control the progress of the teleprompter system document. Speech recognition technology is divided into online speech recognition technology and offline speech recognition technology. Considering the usage environment and cost of the intelligent lectern, offline speech recognition technology is used here [21, 22].

In the traditional conference speech conditions, speakers need to bring their own paper scripts or use the way described in the previous section for automatic page turning, but none of these approaches can achieve automatic recognition of the speaker's real-time speech progress. The current artificial intelligence and speech recognition technology continues to develop, and application popularity for continuous voice recognition application technology has matured. How to more fully combine intelligent video and audio technology and conference speech needs has become the focus of application, and automatic speech recognition

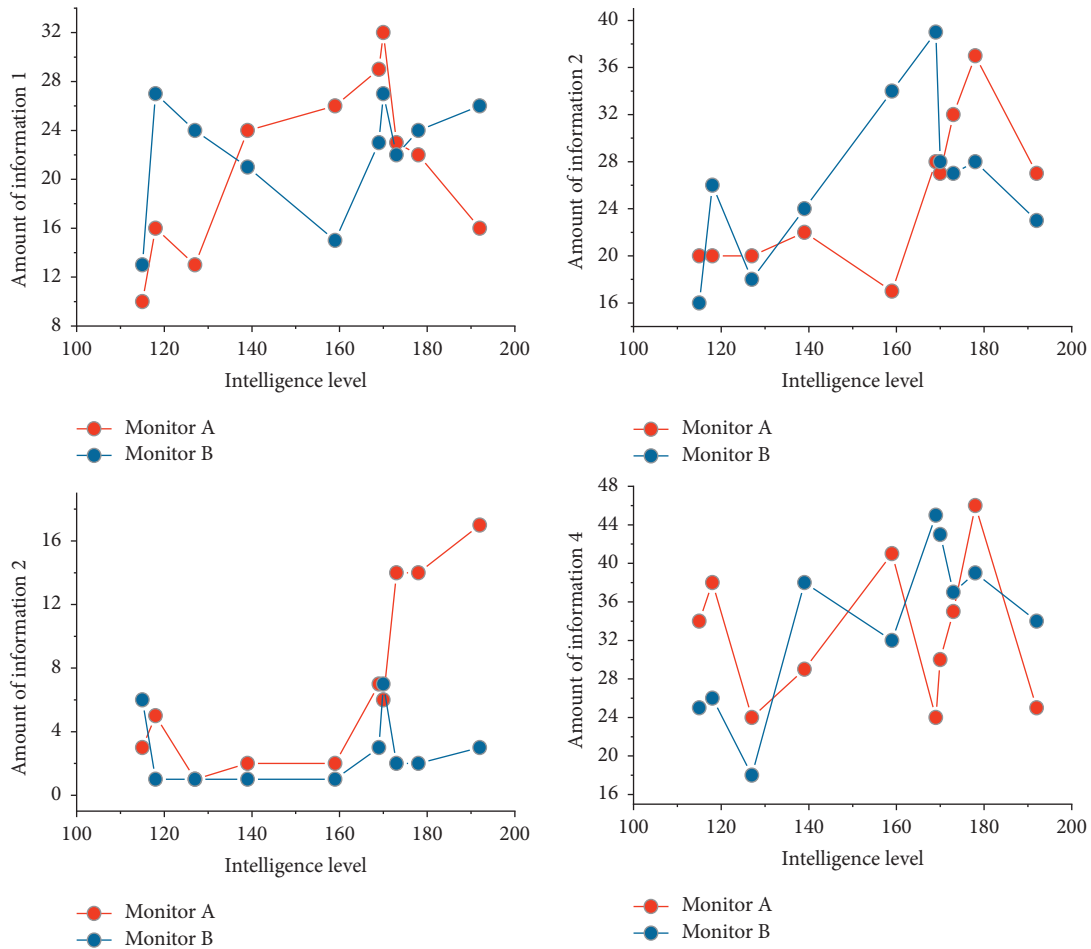


FIGURE 2: Relationship curve between the amount of information received and the product intelligence and personalization.

technology has become the “artificial intelligence + speech” breakthrough. Speech recognition system is divided into three layers: platform capability service, business software application, and speech intermediate control.

The platform capability service layer provides the intelligent speech recognition system server program, WEB server, speech capability platform service engine (speech recognition platform), database management, system resource management, and other related service functions required by the system, and on the premise of completing the basic functions, the application capability can be optimized according to the actual situation of system operation to improve the application level.

The business software application layer provides the information display function of the intelligent speech recognition system used by the speaker and provides the text display corresponding to the real-time transcribed speech and the processing function of various basic document information.

The speech middle control layer mainly provides the speech recognition middleware program, transmits data information with each other with the speech collection equipment picking port and speech recognition SDK interface, completes the functions of speech data collection, processing, storage, and network transmission, and interacts

with the speech capability platform service engine in the platform capability service layer. The intelligent speech recognition system includes speech recognition server, real-time recognition terminal, multichannel speech processor, professional conference microphone, router, and other products, among which the speech recognition server realizes the deployment engine and other types of core capability software, achieves a high degree of equipment integration, reduces capital investment, and provides recording service processing, data transmission, and other capabilities. The analytical formula between key objects is shown in the following equations. According to the formula, we can get the histogram of the relationship between the integration degree of speech recognition equipment and capital investment (three scenarios), as shown in Figure 3. It can be seen from the figure that the higher the integration degree of speech recognition server equipment, the lower the capital investment. The real-time recognition terminal is mainly used to deploy client software and provide the operation of each function of the software. The multichannel voice processor converts the audio data of analog microphone into network data through professional voice acquisition technology, which is used as the data source of the whole system voice. The system topology diagram is shown in Figure 4.

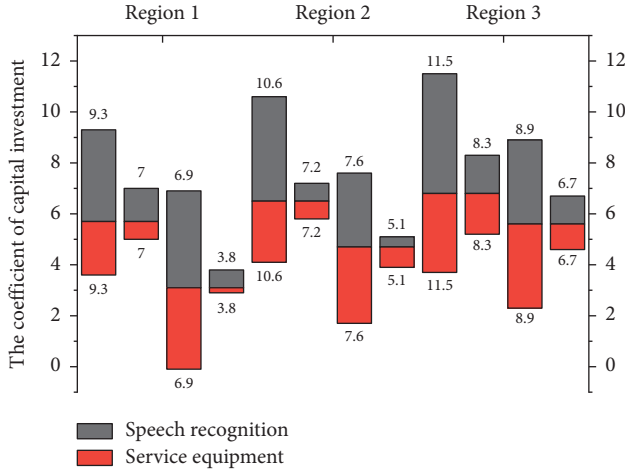


FIGURE 3: Relationship between high integration of equipment and capital investment.

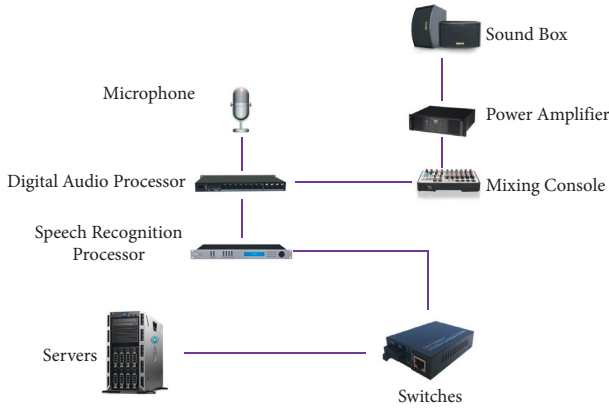


FIGURE 4: Speech recognition system topology diagram.

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2},$$

$$D^n = \sum_{i=1}^j \sum D(X, Y_i^{n-1}), \quad (2)$$

$$\delta^n = \frac{|D^{n-1} - D^n|}{D^n},$$

where  $D$  is the degree of integration of speech recognition service equipment;  $D^n$  is the degree of integration of all equipment;  $\delta^n$  is the coefficient of capital investment.

**2.4. Cloud Computing.** In essence, cloud computing is a virtualized resource. This computing method dynamically provides service expansion through the Internet. It is a pay-per-use model to provide available, convenient, and on-demand network access. Cloud computing is an important research area for future development, in terms of application, it has low requirements for the client devices, and the

devices themselves do not need to be too highly configured because the resources used are from the cloud, and as long as the network is smooth, data and application sharing can be achieved, which can be analyzed through the following formula.

Currently, cloud computing and speech recognition technology have become emerging teaching methods in the education industry [23, 24], and the speech recognition module in the cloud architecture of the intelligent classroom management system can respond to a variety of user requests and can take advantage of a large amount of cloud data to improve the performance of the speech recognition system. Speech cloud uses cloud computing to achieve fast speech applications, which in this system mainly recognizes the human voice. Cloud mode speech recognition and interaction service is a new direction for future research and application. In this regard, the technologies of KDDI, Ali cloud, Baidu, and Tencent cloud are in the leading position in China.

$$X = \{X_1, X_2, \dots, X_m\},$$

$$X_i = \{X_1, X_2, \dots, X_n\}, \quad (3)$$

where  $X$  is the configuration parameters of equipment.

**2.5. Voice Recognition.** Voice recognition technology is mainly divided into two categories, that is, voice meaning recognition and voice similarity recognition. Sound meaning recognition is to transform human voice into text by analyzing the human voice and finding the characteristics of pronunciation from it, which is usually used in such fields as fast inputting information, artificial intelligence, and communication between human and computer through voice. Similarity recognition of voice is to compare the target voice object to be recognized with the voice sample and check whether the similarity between the target voice and the sample can be achieved [25, 26]. The computer and human are basically similar in terms of speech recognition processing. A complete speech recognition system is generally divided into three parts, that is, speech denoise pre-processing and extraction of features of speech, acoustic modeling and pattern matching, and language modeling and language processing. It is in the noisy environment; due to the complexity of the actual environment, noise reduction processing is of great practical significance. For the purpose of lifting up the level of speech denoising and the accuracy of speech recognition system, wavelet denoising technology is often applied to speech recognition. The flow of speech recognition is shown in Figure 5.

### 3. System Design

At present, the speech cloud is widely used in the general field, with a huge amount of user speech data and relatively high accuracy of speech recognition. However, in the field of education, the commands that need to be recognized in the control of multimedia equipment in school smart classrooms are relatively fixed, so the local speech library can be customized to meet the personalized needs of users and

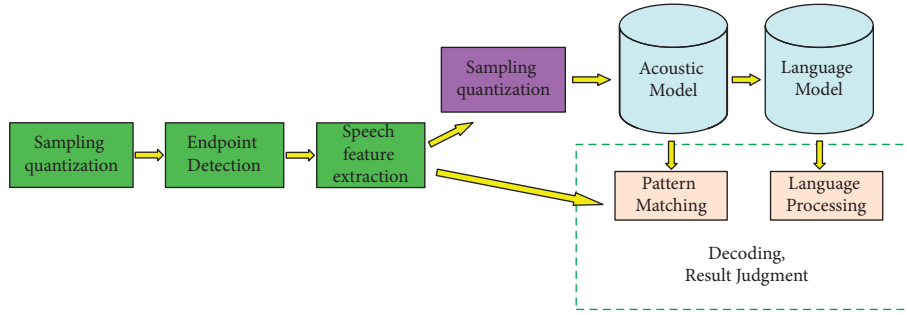


FIGURE 5: Flowchart of speech recognition.

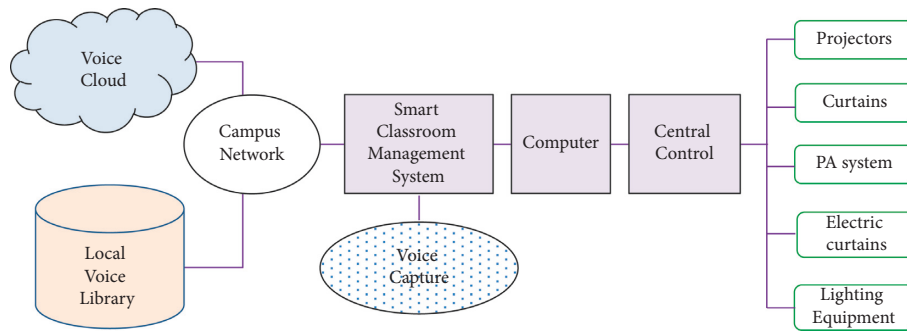


FIGURE 6: System architecture diagram.

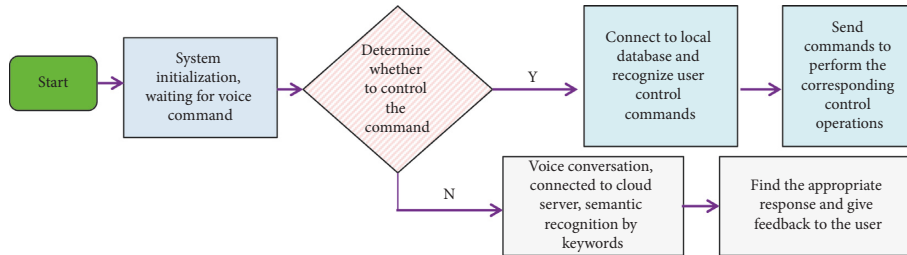


FIGURE 7: System workflow chart.

make up for the shortcomings of the speech cloud such as slow recognition speed due to too wide a search range, heavy reliance on network, multilink leakage, risk concentration, and reduced flexibility of users' control of data and technology.

**3.1. System Architecture.** The overall structure of the speech recognition-based intelligent classroom management system with cloud computing architecture is shown in Figure 6.

**3.2. System Workflow.** Generally, the computer in the smart classroom automatically opens at the set time and starts the client of the management system. First, the software loads the basic syntax package for login, initializes the login speech recognition engine, initializes the interface, and waits for the user to login; then, after successful login, it waits for the user's voice command; after the teacher user issues the correct voice login command, the system starts to judge; if it is a command to control multimedia devices, it connects to the local voice library and controls the devices through the central control serial port after recognition if it belongs to

the general chat. If it belongs to the voice conversation category of general chat, it connects to the voice cloud, finds the answer after recognition, and gives feedback to the user through voice or text. The relationship can be accurately predicted by the following formula. The flowchart of the system is shown in Figure 7.

$$\begin{aligned} \cup Q_j &= O, \\ Q_{j1} \cap Q_{j2} &= \emptyset, \end{aligned} \tag{4}$$

where  $Q$  is the voice login command which the teacher sends to the user;  $1 < j < n$ .

## 4. Practical Application

**4.1. Application of Aliyun.** Teachers and students realize human-computer interaction with machines, which involves human speech recognition and needs to be connected with the speech cloud. At present, the voice recognition interface of KDDI is not free, and the application and approval process of voice recognition of Tencent cloud is relatively



long and tedious. Finally, after comparing the voice clouds of Ali and Baidu, relatively speaking, Ali cloud is easier to use, so the SDK of voice recognition of Ali cloud is used. In addition, the commonly used voice recognition module is FreeSWITCH; the advantages are open-source, cross-platform, scalability, multiprotocol, and so on; it is based on Ali cloud, easy to use, and therefore popular with secondary developers. Its main development language is C, some modules use C++, and it supports SIP, H323, Skype, Google Talk, and many other communication protocols. Voice service Aliyun SDK source code is available on the Github open source platform, and using CommonRequest to invoke the SDK's core library directly is highly useful in development. The process of implementing this function is as follows: firstly, the collected user voice data will be sent to the backend, then the backend will send the received voice input stream to the Ali cloud server side, which will convert the voice into text, and finally, the processed voice data stream will be returned to the frontend.

**4.2. Application of Local Speech Library.** Microsoft Speech SDK is a toolkit launched by Microsoft to develop speech applications and speech engines on Windows platform [27]. It contains various components for speech recognition. There are many examples of secondary development using Microsoft's speech recognition development toolkit, and the methods and ideas from other studies are referenced here.

In order to reference the COM component provided by the SDK, the VisualStudio.NET development platform is used as an example, and the component is referenced by selecting Project|Add Reference in the menu and then clicking on the COM tab and selecting Microsoft Speech Object Library. NET for speech recognition module development mainly uses three APIs: ISpRecognizer interface is responsible for interacting with the underlying RecognitionEngine, which is the speech recognition engine interface; ISpRecoContex interface is responsible for sending and receiving messages, which is the main interface to complete the recognition task; ISpRecoGamma interface is responsible for creating, loading, and activating grammar rules and is the grammarian interface. The Microsoft Speech SDK software development kit provides components for speech recognition, and the system was developed using the C# language, which is inherently well integrated. It is also important to note that since the downloaded SDK only supports English and teachers and students mostly communicate with each other in Chinese, the SDK language package SpeechSDK51LangPack should be downloaded and installed.

**4.3. Serial Port Control.** At present, the centralized control system of multimedia equipment (referred to as central control) on the market is becoming more and more advanced, and some of these smart classroom products are designed to meet the needs of information-based teaching, and they apply broadcast-grade product technology to the campus, leading the new trend of smart teaching. Nevertheless, since the multimedia equipment of each school is

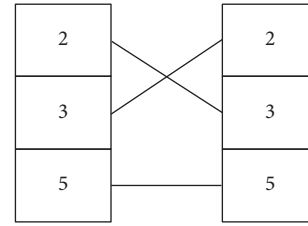


FIGURE 8: Serial control cross wire soldering diagram.

more or less different, the central control may not be able to control some multimedia equipment, so we should redevelop the module with the characteristics of the school according to the actual situation of the school.

Most computers and multimedia equipment have RS-232 interface; if not, you can also convert the USB port to RS-232 interface through the “USB to RS-232” data cable, and then, certain multimedia equipment that the central control cannot directly control can be directly connected to the computer through the device serial port with a network cable by the computer to directly control. The advantage of serial communication is that data can be transmitted over long distances, the use of ordinary network cable soldering is low cost, bandwidth can also fully meet the requirements but also customize the protocol of transmission, and data transmission is more reliable [28]. There are nine pins of the RS-232 interface, of which pin 2 is used to receive data, pin 3 is used to send data, and pin 5 is the signal ground. 9-pin serial port uses only the second, third, and fifth three of these pins to send and receive data, that is, solder wire one end of the serial port in the order of three pins, adjust the other end of the second and third pins a little, the fifth pin remains unchanged, solder wire the other end of the serial port in the order of three pins. The mathematical relationship is as follows. The final crossover line produced is shown in Figure 8.

$$Y = \{Y_1, Y_2, \dots, Y_m\}, \quad (5)$$

$$Y_j = F(X_i),$$

where  $F$  is the data sending and receiving;  $Y$  is the number of serial ports;  $1 < i; j < n$ .

## 5. The Relationship between Smart Classroom and Chinese Teaching

**5.1. Teaching in the Smart Classroom.** Smart classroom is a product of the deep integration of information technology and education teaching, which seamlessly connects teaching and learning inside and outside the classroom and provides a personalized, intelligent, and digital learning environment. The smart classroom environment effectively integrates core functional elements such as teaching resource management, real-time content delivery, learning scenario collection, instant feedback and evaluation, and member interaction and communication by using “cloud service + mobile application” to efficiently organize the three teaching links of teachers and students before, during, and after class.

Smart classroom is student-centered, emphasizing students' autonomous learning and collaborative learning

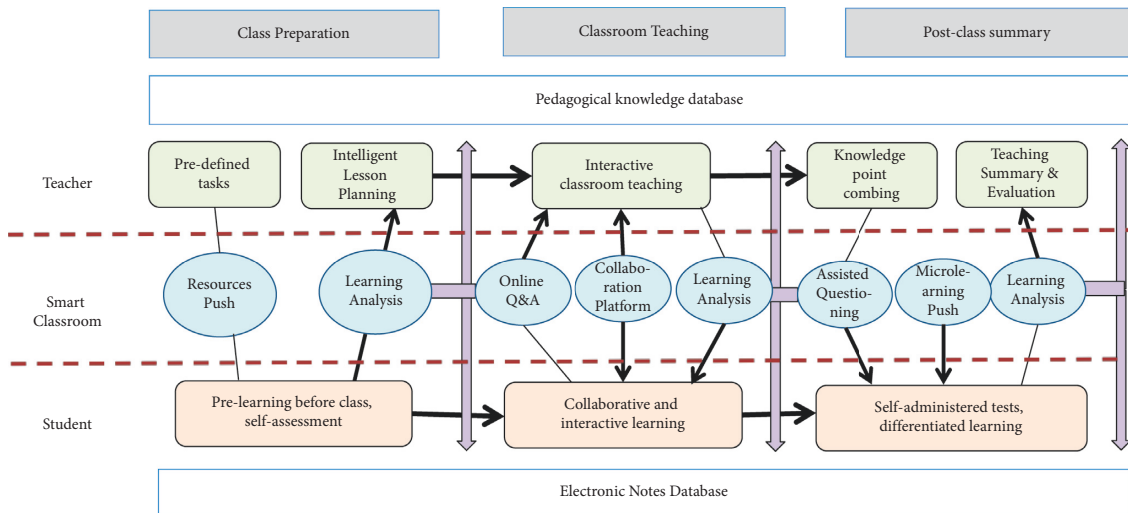


FIGURE 9: Chinese learning model in smart classroom.

among students. Teachers in the smart classroom are more likely to create, collect, and organize educational resources, set goals and assessment criteria, monitor the progress of student-initiated learning in the classroom, and provide timely feedback on questions raised by students. These features of teaching and learning in the smart classroom also fit the requirements of the language teaching process.

**5.2. Characteristics of Chinese Language Teaching and Learning.** First, language teaching emphasizes the learner as the root and stresses that students should be active participants and language constructors in the learning process, and teachers are collaborators and facilitators of learning. Consequently, teachers need to create rich interactive conditions to promote the smooth interaction between teachers and students.

Second, analyzed from the perspective of the language teaching environment, teaching Chinese in the native language environment should try to create a relaxed and natural language atmosphere. Therefore, teachers need to simulate and create language situations and use text, graphics, sound, and video in a comprehensive way to push language information and stimulate learners from multiple senses as much as possible.

Finally, language teaching relies on rich language contexts and the need to obtain timely feedback on learning effects during repeated language training, to achieve interactive communication and personalized learning.

**5.3. The Chinese Learning Model in the Smart Classroom.** In general, the integration of smart classroom and Chinese teaching is an interactive, repeated, and circular process, which runs through the teaching process of preclass preparation, classroom teaching, and postclass summary. This paper moderately simplifies the “universal learning model in the smart classroom” given in the related literature and builds a set of learning models in the smart classroom environment based on the teaching knowledge database and

the electronic notes database, using three times of learning situation analysis as the feedback mechanism (Figure 9). The learning model contains only five teacher activities and three student activities, while the smart classroom platform takes up the tasks of resource pushing, information interaction, and learning situation analysis, making the process of smart teaching more simplified and easy to carry out.

**5.4. Improvement of Chinese Teaching Tasks in the Smart Classroom.** In this paper, we focus on two learning tasks in Chinese teaching, namely, vocabulary and listening, and use the Xunfei Smart Education Platform as a support to explain in detail the improvement methods and technical advantages of the smart classroom for these two teaching tasks.

First, the smart classroom environment can provide students with repeatedly trained vocabulary learning scenarios, push related vocabulary resources, and generate differentiated vocabulary memorization strategies.

Second, the smart classroom environment can provide immersive Chinese listening and speaking contexts, provide students with an interactive and self-directed learning environment before and after class sessions, and use language recognition technology to enhance the effectiveness of listening and speaking interactions in the classroom.

## 6. Optimization of Vocabulary Learning Methods in the Smart Classroom

**6.1. Characteristics of Vocabulary Learning.** Vocabulary is the foundation of Chinese learning and plays a crucial role. However, most vocabulary teaching activities are one-way communication and lack of association, which leads to students’ poor impression of the learned vocabulary and easy forgetting. There are two main reasons for the unsatisfactory effect of vocabulary learning. First, vocabulary teaching does not create relevant context and ignores cultural factors. The second is the lack of effective vocabulary learning strategies, mechanical memory, and ease of forgetting.



6.2. *Vocabulary Learning in the Smart Classroom.* With the learning model of smart classroom, the vocabulary teaching process for teachers and students is concentrated in two stages, before and after class.

6.2.1. *Preclass Preparation Stage.* The teacher gives students clear vocabulary learning tasks, and the resource pushing module of smart classroom pushes the basic word meanings, pronunciation, and lexical properties of vocabulary, as well as the related vocabulary with synonymy, antonymy, multiple meanings, and homophones and homonyms to students. After students finish vocabulary preview, the intelligent classroom provides Dictation Training tools and tests to check students' vocabulary learning effect. Learning analysis can summarize students' questions and guide teachers in preparing lessons intelligently. Smart classroom assists students in preclass prereading and records knowledge points and wrong information in electronic notes.

6.2.2. *Postclass Summary Stage.* Smart Classroom provides a punch card tool to urge students to study words in a planned manner. As the number of vocabulary exercises increases, the granularity of error information in smart classroom's eNotes is refined not only to count which words are error-prone but also to record what types of errors are made (e.g., spelling errors, mispronunciation, word meaning errors, lexical errors, and associated word errors), which can guide students to more scientific vocabulary learning strategies. Differentiated learning strategies will guide the problem setting module and microclass push module of the intelligent classroom platform to help students master error-prone knowledge more accurately and effectively.

## 7. Optimization of Listening and Speaking Training under Smart Classroom

7.1. *Characteristics of Listening and Speaking Training.* The improvement of Chinese listening and speaking ability requires continuous training in language scenarios, and the teaching time of the Chinese classroom is difficult to meet the repeated practice requirements, which leads to the traditional teaching environment where teachers emphasize reading and writing but not listening, and students are unwilling to spend energy on listening and speaking training, making it difficult to improve listening and speaking ability. With the gradual maturity of Xunfei's speech recognition technology and deep learning technology, the classroom integrates independent listening and speaking training scenes and makes use of the openness, sharing, and interaction of the platform to make up for the deficiency of listening and speaking training in actual Chinese teaching.

7.2. *Hearing and Listening Training in the Smart Classroom.* The biggest advantage of smart classroom in improving listening and speaking training is that it provides an immersive language learning environment, which not only

develops students' listening and speaking skills but also improves students' independent learning ability. The listening training platform has a set of training tools for listening, pronunciation, and conversation.

Teachers can use these tools to organize videos and audios related to the course and associate them with a bank of test questions in an easy-to-follow format. Students play the audio and follow along and record it, and the platform uses voice recognition technology to give pronunciation accuracy, which unifies students' listening and speaking skills.

Smart classroom's listening and speaking training platform can also set up common conversational scenarios such as study, life, business, travel, and dating, allowing students to train oral communication skills with the help of human-computer interaction. At the same time, there are also language environments that promote interest in learning English, such as online dubbing of movie clips and MTV of Chinese songs. These environments have a positive effect on enhancing students' interest in listening and speaking.

## 8. Conclusion

Informationization in higher education enters the stage of smart campus, while classroom is the main position of teaching, and the construction of smart classroom is the major trend of future development. After improving and optimizing the construction of digital campus, insisting on building a smart campus with service as the main line, it is the primary task of the school information center to let teachers and students enjoy the convenient effect brought by the information service of the school. Smart classroom realizes active and autonomous learning, audio-visual equipment is intelligent and humanized, and the information services of the Internet and campus network are applied to the field of teaching. In terms of human-computer interaction, the cloud-based approach combined with the local voice library for speech recognition and self-developed management system can save financial and material resources, enhance the security of school data and information, and provide higher flexibility for future upgrades and optimization of the system, as well as improving the development and practical application capabilities of the school's research team. At present, face recognition, as one of the successful applications in the field of image analysis and processing, is gradually integrated into people's lives. Then, the construction of smart campus and smart classroom, in addition to improving the accuracy of voice recognition, should also study the application of face recognition technology in this area in the future so that artificial intelligence technology can bring greater convenience to all aspects of people's lives.

Smart classrooms bring changes to the existing teaching methods and can assist teachers and students in gaining a new learning experience. This paper takes vocabulary learning and listening and speaking training, two tasks suitable for independent learning, as the starting point to show the improvement of the smart classroom for the

Chinese teaching environment and its advantages for enhancing learning interest and accumulating knowledge. The teaching model under smart classroom is a new direction for future teaching development, which needs to be continuously practiced, optimized, and improved by front-line teaching staff.

## Data Availability

The dataset can be accessed upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] S. Al-Sharhan, "14 Smart classrooms in the context of technology-enhanced learning (TEL) environments," *Transforming Education in the Gulf Region: Emerging Learning Technologies and Innovative Pedagogy for the 21st Century*, vol. 188, pp. 1–10, 2016.
- [2] M. Li, "Smart home education and teaching effect of multimedia network teaching platform in piano music education," *International Journal of Smart Home*, vol. 10, no. 11, pp. 119–132, 2016.
- [3] J. Wang, J. Zhang, J. Fan, S. Zhang, J. Wang, and Y. Geng, "Design and application of smart vocational education platform based on new generation information technology," in *Proceedings of the 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, pp. 505–509, IEEE, Chongqing, China, July 2020.
- [4] D. K. Mohanachandran, C. T. Yap, Z. Ismaili, and N. S. Govindarajo, "Smart university and artificial intelligence," *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success*, pp. 255–279, Springer, Cham, Switzerland, 2021.
- [5] J. B. Smart and J. C. Marshall, "Interactions between classroom discourse, teacher questioning, and student cognitive engagement in middle school science," *Journal of Science Teacher Education*, vol. 24, no. 2, pp. 249–267, 2013.
- [6] H. Wang, "Research on the talent cultivation model of the integration of production and education in higher vocational education under the background of "Internet+," in *Proceedings of the 2021 16th International Conference on Computer Science & Education (ICCSE)*, pp. 943–946, IEEE, Lancaster, UK, August 2021.
- [7] A. Fung, H. Russon Gilman, and J. Shkabatur, "Six models for the internet + politics," *International Studies Review*, vol. 15, no. 1, pp. 30–47, 2013.
- [8] S. Mumtaz, "Factors affecting teachers' use of information and communications technology: a review of the literature," *Journal of Information Technology for Teacher Education*, vol. 9, no. 3, pp. 319–342, 2000.
- [9] D. N. E. Phon, M. B. Ali, and N. D. Abd Halim, "Collaborative augmented reality in education: a review," in *Proceedings of the 2014 International Conference on Teaching and Learning in Computing and Engineering*, pp. 78–83, IEEE, Kuching, Malaysia, April 2014.
- [10] W. Li, "Design of smart campus management system based on internet of things technology," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 3159–3168, 2021.
- [11] M. C. Hill and K. K. Epps, "The impact of physical classroom environment on student satisfaction and student evaluation of teaching in the university environment," *Academy of Educational Leadership Journal*, vol. 14, no. 4, p. 65, 2010.
- [12] J. Mohelníková, M. Novotný, and P. Mocová, "Evaluation of school building energy performance and classroom indoor environment," *Energies*, vol. 13, no. 10, p. 2489, 2020.
- [13] S. Y. Phoong, S. W. Phoong, S. Moghavvemi, and A. Sulaiman, "Effect of smart classroom on student achievement at higher education," *Journal of Educational Technology Systems*, vol. 48, no. 2, pp. 291–304, 2019.
- [14] M. K. Saini and N. Goel, "How smart are smart classrooms? A review of smart classroom technologies," *ACM Computing Surveys*, vol. 52, no. 6, pp. 1–28, 2020.
- [15] S. Shaikh Naziya and R. R. Deshmukh, "Speech recognition system—a review," *IOSR Journal of Computer Engineering*, vol. 8, no. 4, pp. 3–8, 2016.
- [16] Z. Bai, G. Sun, H. Zang et al., "Identification technology of grid monitoring alarm event based on natural language processing and deep learning in China," *Energies*, vol. 12, no. 17, p. 3258, 2019.
- [17] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of big data*, vol. 2, no. 1, p. 1, 2015.
- [18] J. Wu, "English real-time speech recognition based on hidden markov and edge computing model," in *Proceedings of the 2021 third international conference on inventive research in computing applications (ICIRCA)*, pp. 376–379, IEEE, Coimbatore, India, September 2021.
- [19] Z. Xiangyan, "Application study of modern educational technology under cloud computing platform," in *Proceedings of the 2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 122–125, IEEE, Macau, China, March 2016.
- [20] M. Kleinert, H. Helmke, S. Moos et al., "Reducing Controller Workload by Automatic Speech Recognition Assisted Radar Label Maintenance," in *Proceedings of the 9th SESAR Innovation Days*, Athens, Greece, 2019.
- [21] L. P. Rondon, L. Babun, K. Akkaya, and A. S. Uluagac, "HDMI-watch: smart intrusion detection system Against HDMI attacks," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2060–2072, 2021.
- [22] R. T. Kreutzer and M. Sirrenberg, *Understanding Artificial Intelligence*, Springer International Publishing, Berlin, Germany, 2020.
- [23] L. Wei, "Study on the application of cloud computing and speech recognition technology in English teaching," *Cluster Computing*, vol. 22, no. S4, pp. 9241–9249, 2019.
- [24] X. Luo and L. Xie, "Research on artificial intelligence-based sharing education in the era of Internet+," in *Proceedings of the 2018 International conference on intelligent transportation, big data & smart city (ICITBS)*, pp. 335–338, IEEE, Xiamen, China, January 2018.
- [25] A. B. Hancock and L. M. Garabedian, "Transgender voice and communication treatment: a retrospective chart review of 25

- cases,” *International Journal of Language & Communication Disorders*, vol. 48, no. 1, pp. 54–65, 2013.
- [26] D. Stowell, M. D. Wood, H. Pamuła, Y. Stylianou, and H. Glotin, “Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge,” *Methods in Ecology and Evolution*, vol. 10, no. 3, pp. 368–380, 2019.
- [27] M. R. Kamarudin, M. Yusof, and H. T. Jaya, “Low cost smart home automation via microsoft speech recognition,” *International Journal of Engineering & Computer Science*, vol. 13, no. 3, pp. 6–11, 2013.
- [28] P. Popovski, Č. Stefanović, J. J. Nielsen et al., “Wireless access in ultra-reliable low-latency communication (URLLC),” *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5783–5801, 2019.