Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 7873715, 10 pages
https://doi.org/10.1155/2023/7873715

Research Article

Design and Application Research of Embedded Voice Teaching System Based on Cloud Computing

Yueying Li and Feng Wu

College of Information Engineering, Xinyang Agriculture and Forestry University, Xinyang 464000, China

Correspondence should be addressed to Feng Wu; wufeng@xyafu.edu.cn

Received 6 February 2023; Revised 22 February 2023; Accepted 27 February 2023; Published 28 April 2023

Academic Editor: Mohsen Ahmadi

Deep changes are occurring in the components and forms of education as a result of the ongoing integration and development of emerging technologies like cloud computing, mobile computing, and artificial intelligence with teaching and learning, and the digital transformation of education is consistently being pushed to new heights. Simultaneously, China's higher education has concurrently reached the stage of popularization. The digitalization of higher education is related to the development quality and value proposition of higher education and determines whether it can adapt to the needs of quality diversification, lifelong learning, training personalization, and governance modernization in the popularization stage. As a result, the current and future phases of China's higher education reform call for accelerating the pace of higher education's digital transformation and guiding the high-quality growth of higher education with digital innovation. The application potential of intelligent learning systems in higher education is becoming more and more clear in this context. In view of this, this work draws from previous research and experiences to build and implement an embedded voice teaching system based on cloud computing and a deep learning model to meet the development needs of the current digital transformation of higher education. On the one hand, the new system can well compensate for the flaws and shortcomings of the current teaching means in universities and realize the accompanying ubiquitous learning by relying on the powerful storage and computing capacity of the cloud computing platform. On the other hand, this study designs a set of voice recognition methods integrating HMM + LSTM to enhance the embedded voice system's recognition performance, ultimately allowing for the voice recognition feature to be implemented in the pedagogical system. When it comes to processing audio signals, the hybrid model makes use of both the HMM's robust time processing capability and the deep neural network's robust characterization capability and generalization performance. As a result, the voice recognition rate, anti-interference performance, and noise robustness can all be significantly improved. Finally, the embedded voice system is put through its paces in an experimental setting to gauge its performance and functionality. The results of the tests demonstrate that the created hybrid model has high recognition accuracy and good noise immunity, which will be utilized as a foundation for the design and development of the final system. Meanwhile, the new system's functional modules have achieved the expected results with good stability and reliability. Trial results gathered through interviews and questionnaires demonstrate that the new system significantly enhances the intelligence and adaptability of college teaching methods and is conducive to promoting the improvement of college students' cultural literacy and innovation ability.

1. Introduction

The power of digital technology accelerates the industrial revolution and technological revolution. As a brand-new method of production, "digital" is gradually altering how people produce, live, think, and perceive the world. Digital transformation is a trend that is permeating a wide range of industries, including social production and economic growth [1]. Digitalization has emerged as the primary force driving education reform in the modern age. The focus of education reform worldwide is now on utilizing new technologies like artificial intelligence and cloud computing to accelerate educational growth. Accordingly, the knowledge, skills, and capacities of the labor force should also change

as a result of the development of digital industrialization and industry digitalization in the digital era, which has led to new demands for the quality of the labor force and talent demand. A more fundamental change is required in higher education, which produces labor for the market. Therefore, the digital transformation of higher education has been given higher expectations and has become a prominent issue in the reform and development of higher education during the epidemic period. In short, the digital transformation of higher education is aimed at reshaping its new capabilities based on the new generation of digital technology, which is related to the development quality of higher education. However, the teaching mode of classroom oriented and teacher oriented is still widely adopted in higher education, where the classroom and the teachers themselves serve as the main focus. There is little opportunity for students to ask questions or share their perspectives, making it difficult to spark their interest in learning. Therefore, it is easy to lead to the poor learning effect of students and then lead to the failure of higher education digitalization to achieve the expected purpose. In response to this problem, accelerating the pace of digital transformation of higher education; vigorously promoting the informatization of higher education, the digitalization of educational resources, the empowerment of educational technology, and the innovation of educational methods; and leading the high-quality development of higher education with digital innovation have become the requirements of our country's higher education reform at present and in the future. With the continuous development of emerging information technologies such as artificial intelligence [2], cloud computing [3], and mobile computing [4] in the field of education and teaching, profound changes are taking place in the elements and forms of education. Therefore, in the era of intelligent big data, how to use the advantages of these technologies to help the digital transformation of higher education is an urgent issue to consider.

As information technology and embedded systems continue to advance at a rapid pace, the integration of voice interaction technologies into embedded systems is a rapidly expanding area of study. In addition, with the wide application of deep learning and cloud computing in voice interaction, voice interaction services in the form of cloud are familiar to the public and gradually applied to real life [5, 6]. Therefore, the integration of embedded voice teaching system into the currently hot cloud computing technology can provide a new way of voice service [7]. On this configuration, all of the processing for the system's voice recognition and synthesis is done remotely in the cloud. In doing so, it is able to provide users with voice interaction services in the cloud, which makes up for the drawbacks of earlier voice technologies. It can reduce the resource overhead while users get the voice service. Moreover, it can also provide personalized services to users according to their needs.

It is for this reason that research into the design of the embedded human-computer speech interaction system based on cloud services and its related technologies is both theoretically and practically significant. Companies like iFLYTEK, which specialize in voice recognition, have developed their own proprietary voice interaction and recognition systems. At present, there are also many APPs developed based on embedded voice recognition technology and cloud computing technology on cell phones to assist students in English learning, which will become an inevitable trend in the market in the future. Related deep learning strategies have also been proposed anew as educational theory research has expanded [8, 9]. The technology was then successfully used by researchers to the field of voice recognition. In the literature [10], good results in big vocabulary speech recognition were achieved by applying deep information networks as a pretraining step for deep neural networks and using a DNN-MM hybrid network model to train acoustic models. According to the literature [11], a recurrent neural network LSTM can be used to learn the context of English sentences in order to build a translation model called a Transformer. The results of the experiments demonstrate that the model may be trained using the input utterance, resulting in an enhanced translation effect.

In conclusion, it has become increasingly common to make use of deep learning and cloud computing to process the voice signals in embedded voice systems as a result of their increasing sophistication and widespread adoption. Therefore, the research and design of an embedded voice teaching system based on cloud computing have opened up a new way to innovate the teaching mode of universities and further promote the digital transformation of higher education. To this end, this study combines the above experience to construct a new embedded voice interaction system by combining cloud computing and voice recognition technology to improve the teaching effect of universities. The following are some of the significant innovations made in this research: to begin, we suggest combining the HMM and a deep neural network to create a hybrid model for voice recognition. The recognition impact can be enhanced by combining the strengths of the HMM's robust timing processing capacity with the deep neural network's robust characterization ability and generalization performance during the processing of the speech signal. Second, the cloud computing platform is used for voice recognition and synthesis in the new system, which has the benefits of low cost, good scalability, large scale, and strong computing power, and thus contributes to the realization of "ubiquitous learning" for the course teaching in universities. Third, the new system gives teachers and students access to a cutting-edge teaching platform that facilitates the sharing of ideas and resources, facilitates student-teacher interaction, and responds to the growing need for a wide range of instructional approaches. Finally, the experimental environments were used to assess the performance and functionality of the embedded voice system. Results from testing indicate that the new system performs as intended in design. It helps boost the improvement of college students' cultural literacy and considerably enhances the intelligence and adaptability of the teaching methods in universities.

2. Related Concepts and Methods

2.1. Voice Recognition Technology. Broadly speaking, the term "voice recognition technology" encompasses both

semantic and vocal recognition. In a narrow sense, it refers to the understanding recognition of speech semantics, also known as automatic speech recognition (ASR). To realize the control of voice to the machine, voice recognition technology takes voice as the research object and then automatically converts the input voice signal into the matching text or command using computer and software [12]. Although numerous approaches have been offered by various researchers towards the improvement of voice recognition technology, the fundamental concepts remain mostly unchanged. Figure 1 basically depicts the recognition principles applied by voice recognition systems during the processing of voice data.

As voice recognition technology continues to evolve, many learning systems based on this technology are being employed to address the demand for personalized practice and real-time feedback for students.

2.2. Basic Methods

2.2.1. HMM. The hidden Markov model, or HMM, is considered to be one of the most representative models in the field of voice recognition [13]. It refers to the process of transferring one state to another state by means of a hidden form quite frequently. This model is a mathematical and statistical strategy, and the parameters of the hidden state can be discovered by calculating the probability of the states being transferred. It possesses advantageous characteristics such as high operating efficiency and stable operation. Its training process mainly includes forward algorithm, Viterbi, and Baum-Welch [14]. The mathematical description of it is as follows.

Assuming that the HMM is a five-element array (S, P, U, W, and C), its mathematical expression can be described by the following formula:

$$H = (S, P, U, W, C), \tag{1}$$

where S represents the set of states and $P = \{p_{i,j}\}$ denotes the probability of transferring from state i to state j, which can be defined in the following formula:

$$\sum_{N=j=1} p_{i,j} = 1, \quad 1 \le i \le N. \tag{2}$$

$U = \{F[S_1(1)]\}$ in which when $t = 1$, it is the probability of being in state S_i.

W denotes the discrete vocabulary list, which can be defined in the following formula:

$$W = \{W_1, W_2, \cdots, W_m\}, \tag{3}$$

where m represents the number of symbols in W. Suppose $C = \{c_j(W_k)\}$ denotes the probability of symbol W_k under state S_j, which is defined in detail in the following formula:

$$\sum_{k=1}^{M} c_j(W_k) = 1, \quad 1 \le j \le N. \tag{4}$$

2.2.2. LSTM. As one of the classical models in recurrent neural networks, the long short-term memory (LSTM) neural network is mainly evolved from the RNN [15]. It has two transfer states. LSTM is capable of learning long-term dependant information while processing current input, which allows it to efficiently solve the problems of gradient expansion and gradient disappearance that are present in RNN. The network may be broken down into its primary components, which are the input gate, the output gate, and the forgetting gate [16]. The structure of the network can be seen in Figure 2.

The input to the LSTM cell usually includes three forms: current moment input, previous moment hidden layer state, and cell state, which are represented by three variables x_t, h_{t-1}, and c_{t-1}, respectively. σ denotes the sigmoid activation function. In the LSTM model, three gating mechanisms, namely, input gate i_t, output gate o_t, and forgetting gate f_t, are responsible for filtering and updating the input information, and their expressions are shown in formula (5), formula (6), and formula (7), respectively.

$$i_t = \delta(W_i[y_{t-1}, x_t] + b_i), \tag{5}$$

$$o_t = \delta(W_o[y_{t-1}, x_t] + b_o), \tag{6}$$

$$f_t = \delta\left(W_f[y_{t-1}, x_t] + b_f\right), \tag{7}$$

where W_t, W_o, and W_f denote the connection weights, respectively; b denotes the bias matrix; and y_{t-1} denotes the output at moment $t - 1$.

In addition, the cell state z_t and the hidden layer state h_t in the LSTM model are shown in formula (8) and formula (9), respectively.

$$z_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{z}_t, \tag{8}$$

$$h_t = o_t \cdot \tanh(c_t). \tag{9}$$

3. Design of Embedded Voice System

The voice training system has evolved to include embedded voice terminals as a result of the rapid growth of embedded technology. This study combines cloud computing, a deep neural network, and an embedded system to create a voice interaction system, which is then applied to the daily teaching and practice of university courses. This is done to meet the personalized and diversified learning needs of students and effectively improve their learning efficiency.

3.1. General Design of the System. By shifting much of the computing and information processing to the cloud, the embedded voice system developed in this study considerably reduces the hardware resources consumed by the client. In this way, each student can learn and practice their courses whenever and wherever they have access to mobile terminals, thus greatly improving the learning efficiency. Concurrently, the system employs intelligent identification technology to deliver intelligent feedback to the learning process, allowing students to learn autonomously even without the assistance of teachers and allowing for greater

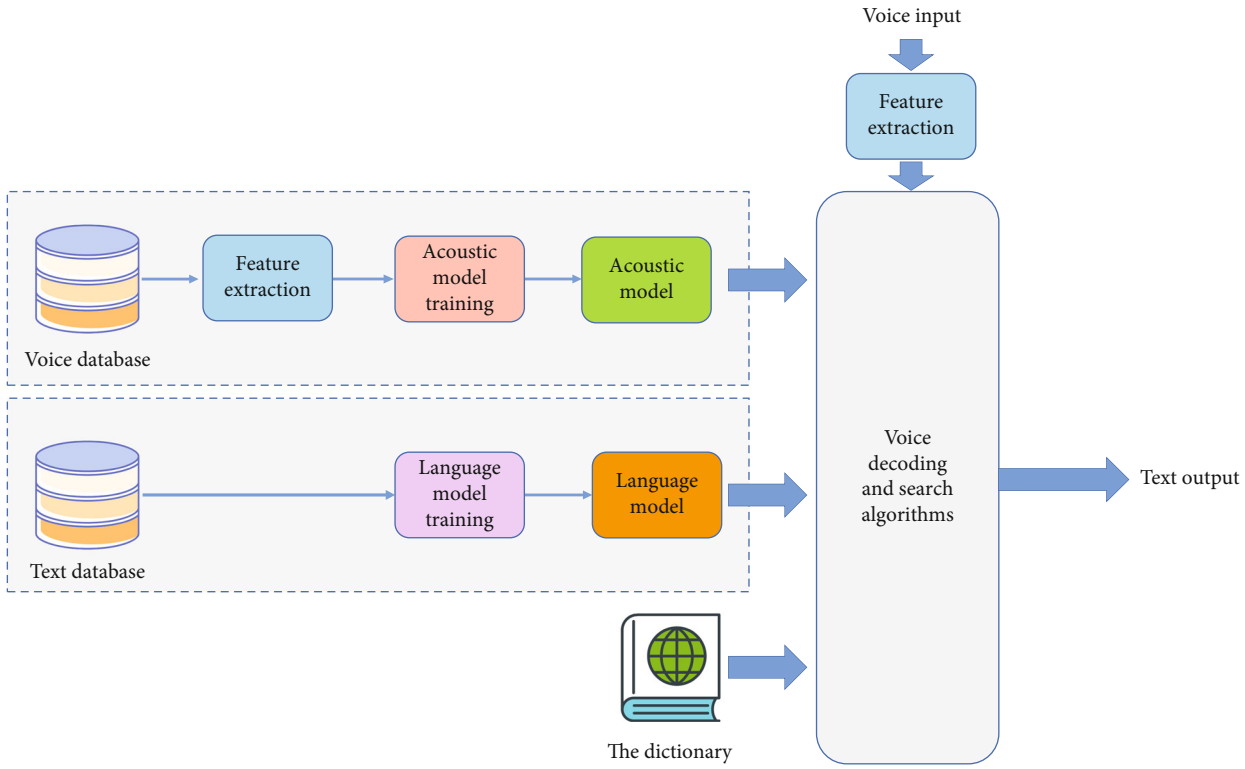

Figure 1: Basic principle of voice recognition system.

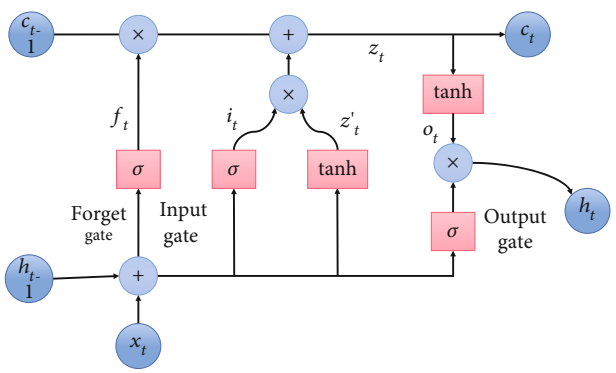

Figure 2: Cell structure of LSTM.

portability in terms of time and location of classroom instruction. Relying on the powerful storage and computing capacity of the cloud computing platform, more advanced voice recognition technologies can be applied to the platform, making the university courses richer and more interesting.

This system allows a single teacher to teach a class of students using only their voice by connecting a teacher's computer to a number of student terminals through the internet and a cloud computing platform. Because the system relies on packet-based data grouping interchange, it is able to implement both voice-based group instruction and interactive instruction [17]. The new system's primary features are as follows: first, the teacher can play a variety of audio files to the entire class for broadcast teaching. Second, students have the option of borrowing audio files from the teacher's computer to use during independent study time. Since this system relies on the powerful storage and computing capabilities of the cloud computing platform, the transmission loss is small, the speed is fast, and the voice is clear and pure, which effectively alleviates the problem of noise interference under the traditional analog technology. It is possible for the teacher and students to converse simultaneously without interfering with one another. Additionally, the system offers a good human-computer dialogue interface, enabling virtually error-free communication between the teacher and students as well as between students.

3.2. Operation Mode of the System. The primary goal of the system applied to university courses is to realize the regular teaching tasks, and the teachers complete the teaching tasks in the mode of centralized lectures or personalized lectures, respectively, while the student terminals are the receivers for listening and learning. In this system, the teacher terminal is generally designed with traditional PC, and its primary purpose is to carry out the regular teaching tasks, which are divided into centralized lecture module and personalized lecture module. In addition, the teacher's terminal adds the function of student learning evaluation, so that the teacher can understand the students' learning situation dynamically and the students can recognize their own learning situation in time. The student terminal is an embedded digital voice terminal designed to receive and play digital voice files or to play voice files in the library on demand according to their existing knowledge or interests. Figure 3 depicts the system's teacher and student terminals in their operation modes.

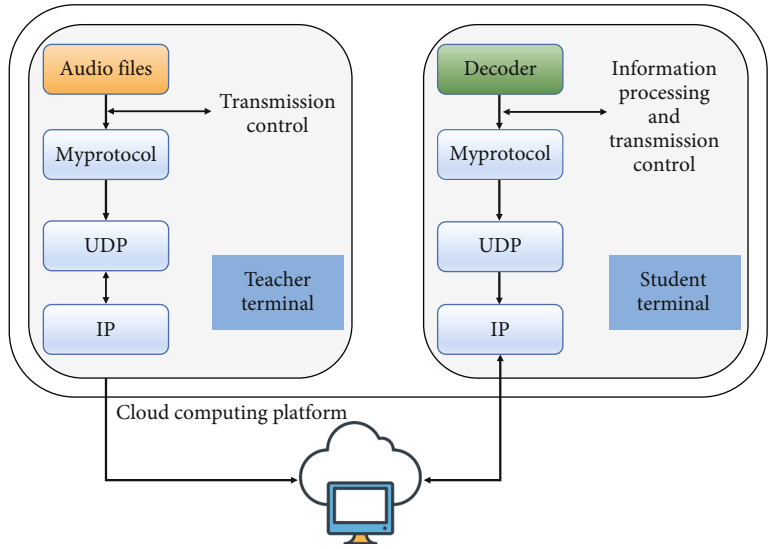

Figure 3: Operation mode of the system.

3.3. Terminal Design of the System

3.3.1. Design of Teacher Machine Terminal. The teacher machine terminal uses a traditional PC with Windows as the operating system and is developed using a high-level programming language. When compared with other computing options, the PC is far superior in terms of speed, memory capacity, and storage. The teacher machine's primary features include identity authentication, roster management, centralized teaching, personalized teaching, assessment of students' learning outcomes, and management of vocal resources. Figure 4 illustrates the teacher terminal's primary features.

The teacher terminal's lecture function flow gives the teacher the option of using either a centralized or individualized lecture format, depending on the needs of the class. In centralized lectures, the teacher broadcasts audio files to all of the students. For personalized lectures, the lecturer might make it more interactive by asking the class questions related to the material being covered.

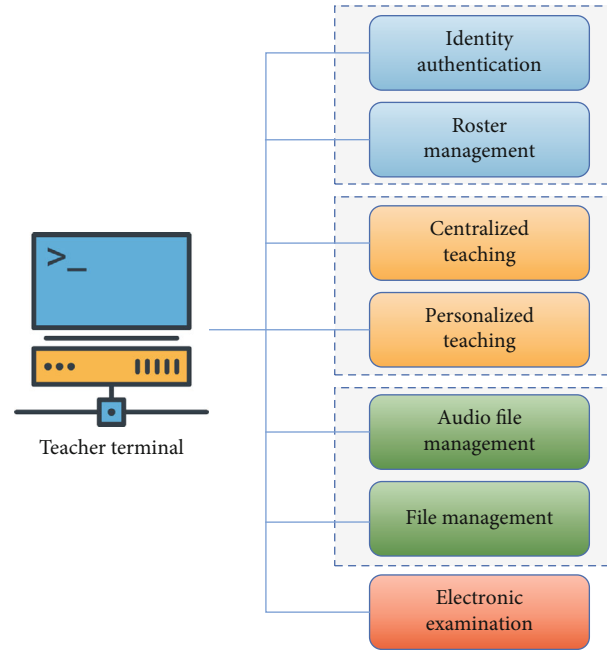

Figure 4: Functional design of the teacher terminal.

3.3.2. The Design of a Student Terminal with an Embedded System. The student terminal in the system adopts the embedded-based terminal design mode, which mainly completes the playback of audio data stream, text display, and keyboard control. The success or failure of the entire embedded voice teaching system hinges on the design of this component. The student terminal's overall design can be seen in Figure 5.

Figure 5 shows that the four primary components of the student terminal are the hardware layer, the operating system layer, the driver layer, and the application layer. This is because the student terminal was developed using the embedded manner. The hardware layer is the hardware platform on which the whole system and applications run, and different hardware environments often need to be configured for different applications. The primary operations of the system, such as task scheduling and control of embedded applications, are realized by means of the operating system layer. Each module's driver design is finished in the driver layer. The application layer is mainly used to analyze the embedded Internet protocol stack, realize the development of the playback program, provide a friendly graphical user interface, and complete the functions of the digital voice classroom. It is worth mentioning that in the teaching mode, we use buffering for student terminals to ensure the flow and continuity of voice in the system.

3.4. Design of a Voice Recognition Algorithm in the System. The aim of the embedded voice system is to use the HMM and the LSTM network model in Section 2.2 to implement

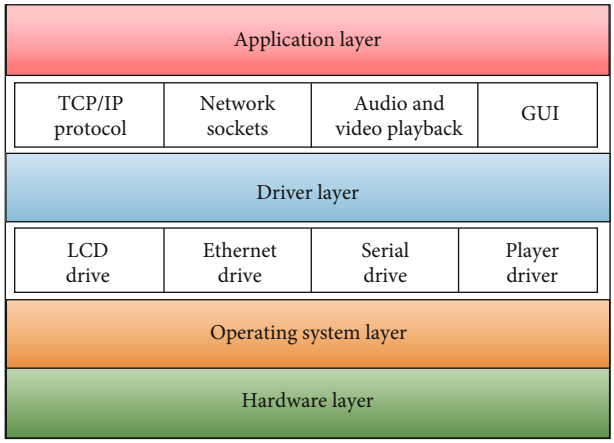

Figure 5: Structure design of student terminal.

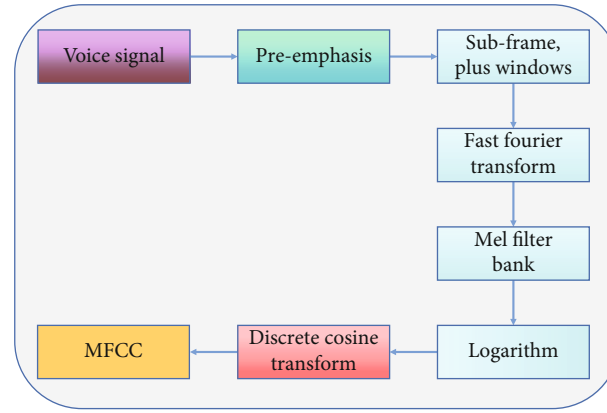

Figure 6: MFCC flow chart.

cloud-based voice recognition and synthesis on the embedded platform to achieve human-computer interaction. Therefore, we need to further design the voice recognition algorithms for the teacher and student terminals in the system.

It is the job of the translator's voice recognition module to first extract features from the voice data and then use those features to translate the voice signal into a text form or command. MFCC (Mel Frequency Cepstral Coefficient) has the advantages of having strong anti-interference and high robustness, and it is easy to achieve a relatively high recognition rate in feature extraction [18]. Therefore, MFCC is selected for feature extraction of voice signals in the embedded voice system designed in this paper. Figure 6 illustrates its specific workflow.

Afterwards, the input voice translation results are utilized to train a model for voice recognition using the HMM approach described in Section 2.2. The model's probability value can be used to determine the optimal state sequence for voice synthesis, and the model's output can then be applied to generate a convincing synthetic voice signal [19].

Last but not least, the LSTM network model is utilized to build a recurrent neural network-based Transformer model from the voice recognition data. An "encoder-decoder" structure is utilized to construct the model. Only a feedforward neural network and a multihead attention mechanism are employed in the encoder and decoder built within the system. Along with this, the text recognition accuracy is enhanced by adding a lexical information vector.

4. System Simulation and Testing

Guided by cloud computing and deep neural network theory, this study relies on the embedded development platform to construct an embedded voice teaching system. Thus, with the purpose to ensure the stability and practicality of the system's application in university course teaching, we conducted a series of system simulation test experiments. First of all, one of the core tasks of the platform is voice recognition with the help of HMM and LSTM network model, so it is necessary for us to test the recognition rate of the system accordingly, which is the basis for carrying out other tests. Secondly, we also conducted some self-tests and questionnaires to check the usability and practicality of the system in order to bring teachers and students a better experience of using the system.

4.1. Data for Voice Recognition. We conducted the appropriate experimental validation to confirm the efficacy of the HMM and the LSTM network model in voice recognition. The key components of the testing setup mainly include hardware devices such as computer and sound card, Win10 operating system, and Matlab experimental programming platform. The voice is sampled at 8 kHz in mono, coded with 16 bits, and saved as a WAV file.

Eight different Chinese commands—"open," "close," "forward," "backward," "zoom in," "zoom out," "play," and "pause"—were used as voice samples for the evaluation. All the voice data for this investigation came from individual recordings made in a specially chosen, less noisy room for the capture of speech signals. A total of 400 sets of voice data were obtained from 10 experimental individuals (5 male and 5 female). 200 data were utilized to train an acoustic model, and the other 200 data were used as samples for testing the effect of voice recognition methods.

4.2. Evaluation of the system's Capacity. The recognition rate measures how well the system understands the user's spoken commands. It is one of the key indicators to measure the system performance. With the purpose to test the HMM-LSTM model's efficacy in voice recognition, we conducted a comparison experiment with the recognition performance of existing HMM and PNN models [20]. Table 1 displays the comprehensive comparative results. In the table, the numbers on the left represent the number of correctly recognized words, while the numbers on the right represent the total number of words recognized. For instance, if 30 samples are input and 28 are correctly recognized, the ratio would be 28/30.

The testing results in Table 1 show that all three models achieve voice recognition rates of 90% or above. Meanwhile, the HMM + LSTM model developed in this research achieves a remarkable 96.25% accuracy in voice recognition.

Table 1: Voice recognition rate of each model in pure voice environment.

Command phrase	HMM	PNN model	HMM + LSTM model
Open	28/30	28/30	29/30
Close	29/30	27/30	29/30
Forward	27/30	27/30	29/30
Backward	28/30	29/30	28/30
Zoom in	29/30	27/30	28/30
Zoom out	28/30	27/30	30/30
Play	28/30	27/30	29/30
Pause	29/30	28/30	29/30
Average recognition rate	94.17%	91.67%	96.25%

The HMM achieves a recognition rate of 94.17%, placing it in the middle of the HMM + LSTM model and the PNN model. Based on a comparison of the three models' recognition outcomes, we can deduce that the voice recognition model using deep neural network technology considerably enhances the performance of the embedded voice system. In addition, we also found that the recognition results based on the HMM + LSTM model also outperformed both the PNN and HMM for the eight voice command phrases in the test experiments. This also shows that the hybrid model can effectively implement its relevant experimental features and can effectively schedule the commands in the system.

We also simulated tests of the noise immunity of the three models to further highlight the benefits and performance of the HMM + LSTM model suggested in this study for voice recognition. In the experiments, we added six different Gaussian noises to the above eight command Chinese phrases for the experiments, and the SNRs were 5 db, 10 db, 15 db, 20 db, 25 db, and 30 db, respectively. Next, voice recognition was performed on samples of each model in the noisy environment. Table 2 displays the experimental results.

Analysis of the experimental data in Table 2 shows that as signal-to-noise ratio drops, the three models' voice recognition rates drop continually. Both the HMM and the PNN model suffer a considerable drop in voice recognition quality. While the recognition rate does drop for the HMM + LSTM model, it does so at a far slower rate than either of the other two models. This reflects to a certain extent that the HMM + LSTM has good anti-interference ability and is robust to noise. This is because the new model combines the powerful timing processing ability of HMM and the superior self-learning and classification ability of the LSTM network model, which can describe the semantic content of speech in more detail, so as to better improve the speech recognition rate, anti-interference performance, and robustness to noise.

4.3. Functional Self-Test of the System. The system's primary goal is to supply students with various educational tools, specifically, voice teaching videos and exercise assessments. As soon as we finished designing and developing the embedded voice system, we began testing its many functional modules in the platform to identify any issues or weaknesses in the system so that we could optimize them and improve the overall user experience as soon as possible. The following is the test for the functional modules of the system.

Table 2: Voice recognition rate of each model under different SNR noise environments.

SNR	HMM	PNN model	HMM + LSTM model
5 db	34.72%	46.85%	59.25%
10 db	61.86%	67.68%	79.43%
15 db	72.39%	75.65%	87.75%
20 db	87.47%	87.26%	92.17%
25 db	89.75%	88.85%	93.25%
30 db	93.15%	90.76%	96.00%

To begin, the various functional modules of the platform involve exercises for practice. For this reason, verifying the accuracy of the system and its analysis of the students' responses is crucial. Functional tests were run on the system to ensure that the exercises were performing as expected. Table 3 displays the detailed results. Secondly, in addition to exercises, the system also provides a variety of learning methods such as audio teaching. It is the feature of the system and one of the important functions of the system. In this function module, it mainly includes independent learning methods such as videos, lesson plans, and exercises. In order to test whether the video playback function is normal, the following tests are conducted, and the detailed test results are shown in Table 4.

Analyzing the test results in Tables 3 and 4, we can see that after several iterations, each functional module of the embedded voice system designed in this research has achieved the expected results, with good stability and reliability, and can get a good user experience.

4.4. Trial of the System. Finally, we conducted a trial experience test on the embedded voice teaching system based on cloud computing to validate its efficacy and feasibility in university curriculum education. One hundred university students were chosen as participants, and they were observed for two weeks. Everyone in the experiment used this system to study for a total of 30 minutes per day. Test information was gathered through interviews and questionnaires after the experiment was over. Figure 7 displays the statistical findings from the survey data.

The statistical results displayed in Figure 7 show that more than 85% of students believe the system is helpful for their course learning and that most students believe that the various learning methods, such as videos and exercises, provided in the system can help them well in their professional knowledge. Especially in the exercise practice, the system can give timely feedback after diagnosis. This interactive way is very helpful for college students' learning. Besides, students have good satisfaction with the interface, operation, and functions of the platform.

In conclusion, we focus on the voice recognition performance of the embedded voice system and some core

Table 3: Exercise function test.

Test 1: exercise function test	
Descriptions	The students carry out exercises in the corresponding function module of the system to test whether its function is realized.
Test plan	Students do exercises, choose answers to look at the analysis.
Data input	According to the prompts given and the corresponding requirements, choose the appropriate answer, and finally, check the analysis and submit.
The expected results	The system will judge the answers chosen by students and give corresponding hints.
The actual results	The actual test results are in line with expectations.

Table 4: Phonetic teaching test.

Test 2: phonetic teaching test	
Description	Students do video learning and use the corresponding functions.
Test plan	Students play instructional videos and add comments below them. They participate in the study, review the lesson plan, and consolidate the learning effect. Finally, they finish the exercises.
Data input	Watch the video explanation, check the lesson plan, and complete the exercises.
The expected results	The video plays smoothly, the comments are displayed correctly, and the problem analysis is provided in time.
The actual results	The actual test results are in line with expectations.

The problem	The statistical results				
Whether the system is helpful for your course learning?	Very helpful	Some helpful	Helpful	General	Without help
	8%	35%	42%	10%	5%
Whether the arrangement of exercises in the system, the degree of examination of knowledge is reasonable.	Very good	Good	General	Bad	Very bad
	18%	60%	12%	6%	4%
Whether the problem diagnosis and feedback provided in the system are helpful to your study?	Very helpful	Some helpful	Helpful	General	Without help
	15%	32%	36%	12%	5%
Whether the presentation of teaching video in the system, the explanation of knowledge points is reasonable.	Very good	Good	General	Bad	Very bad
	30%	43%	13%	8%	6%
Whether the voice teaching video in the system is helpful to your course knowledge?	Very helpful	Some helpful	Helpful	General	Without help
	10%	58%	23%	6%	3%
Are you satisfied with the function of the system?	Very satisfied	Some satisfied	Satisfied	General	Not satisfied
	6%	62%	18%	8%	6%
Are you satisfied with the interface and operation of the system?	Very satisfied	Some satisfied	Satisfied	General	Not satisfied
	5%	35%	52%	6%	2%

Figure 7: Results of the questionnaire survey.

functional models of the system for simulation testing. The results of the tests demonstrate that the voice recognition model developed in this study has high recognition accuracy, good anti-interference ability, and high noise robustness even in noisy surroundings, thanks to the use of deep neural network technology. Therefore, the new model can play a very important role in the noisy laboratory teaching as well. Finally, we verified the effectiveness and feasibility of the embedded voice system in university course teaching by means of self-test experiments and questionnaire surveys.

5. Conclusion

Higher education has been infiltrated by the technological progress brought by the Industrial Revolution 4.0, which forces higher education to face the digital transformation

in all aspects. To this end, exploring the digital transformation of higher education has become an emerging field and has attracted extensive attention of scholars. Meanwhile, with the popularization and application of cloud computing, mobile computing, and deep learning, the new generation of high technology is developing very rapidly, which is also overturning the traditional teaching mode of universities. Digital transformation of higher education can take a step forward with the help of cloud computing, embedded voice systems, and interdisciplinary collaboration. This work proposes and develops an embedded voice interaction system for university course teaching, making use of the enormous storage and computing capability of a cloud computing platform and the distinctive performance of an HMM + LSTM model in voice recognition. Experimental results demonstrate that the new model developed in this study not only achieves higher recognition accuracy than the conventional HMM and PNN voice recognition models but also exhibits strong noise immunity. In addition, the results of the tests conducted on the system's functional modules demonstrate that the new system is stable and reliable. Finally, the system trial results indicate that the new system makes professional course learning of college students more rich and fascinating and helps to increase their learning effect.

While previous work has yielded generally positive results, the embedded voice teaching system developed in this study offers novel approaches for university course teaching in the age of artificial intelligence. Due to the limited ability of individuals, there are still some points in the system that need to be further developed and improved. To begin with, the system's interactivity and interface need to be further streamlined and upgraded to provide users with a better experience. Secondly, the system's functionality should be fine-tuned and its learning resources should be enlarged so that it has greater practical value and more loyal users. Finally, new technologies like big data and edge computing can be incorporated to make personalized recommendations for students and improve their learning efficiency.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This study is sponsored by Xinyang Agriculture and Forestry University.

References

[1] S. Kraus, S. Durst, J. J. Ferreira, P. Veiga, N. Kailer, and A. Weinmann, "Digital transformation in business and management research: an overview of the current status quo," *International Journal of Information Management*, vol. 63, article 102466, 2022.

[2] M. Furini, O. Gaggi, S. Mirri et al., "Digital twins and artificial intelligence," *Communications of the ACM*, vol. 65, no. 4, pp. 98–104, 2022.

[3] J. Liu, J. Wu, and L. Guo, "Construction of emergency dispatching and controlling platform for multi-elevator in cloud computing," *International Journal of Internet Protocol Technology*, vol. 14, no. 4, pp. 232–239, 2021.

[4] T. Wang, B. Lu, W. Wang, W. Wei, X. Yuan, and J. Li, "Reinforcement learning-based optimization for mobile edge computing scheduling game," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, pp. 55–64, 2022.

[5] L. A. Kumar, D. K. Renuka, S. L. Rose, and I. M. Wartana, "Deep learning based assistive technology on audio visual speech recognition for hearing impaired," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 24–30, 2022.

[6] M. Vashishtha, P. Chouksey, D. S. Rajput et al., "Security and detection mechanism in IoT-based cloud computing using hybrid approach," *International Journal of Internet Technology and Secured Transactions*, vol. 11, no. 5/6, pp. 436–451, 2021.

[7] Y. Y. Lin, M. C. Wei, C. C. Sun, W. K. Kuo, F. C. Chan, and Y. C. Liu, "Millimeter wave radar combines long short-term memory and energy storage embedded system for on-street parking space prediction," *Sensors and Materials*, vol. 34, 4 Part 2, pp. 1401–1417, 2022.

[8] T. Wang, J. Li, W. Wei, W. Wang, and K. Fang, "Deep-learning-based weak electromagnetic intrusion detection method for zero touch networks on industrial IoT," *IEEE Network*, vol. 36, no. 6, pp. 236–242, 2022.

[9] T. Wang, K. Fang, W. Wei, J. Tian, Y. Pan, and J. Li, "Microcontroller unit chip temperature fingerprint informed machine learning for IIoT intrusion detection," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2219–2227, 2022.

[10] A. Kumar and R. Aggarwal, "An investigation of multilingual TDNN-BLSTM acoustic modeling for Hindi speech recognition," *International Journal of Sensors, Wireless Communication and Control*, vol. 12, no. 1, pp. 19–31, 2022.

[11] C. Shi, "Research on intelligent language translation system based on deep learning algorithm," *Soft Computing*, vol. 26, no. 16, pp. 7509–7518, 2022.

[12] D. C. Silpani, K. Suematsu, and K. Yoshida, "A feasibility study on hand gesture intention interpretation based on gesture detection and speech recognition," *Journal of Advanced Computatioanl Intelligence and Intelligent Informatics*, vol. 26, no. 3, pp. 375–381, 2022.

[13] S. Rezaei and M. Mehrara, "Dynamics of symmetric informed trading and order flow shock at Tehran exchange stock: a hidden Markov model approach," *Quarterly Journal of Applied Theories of Economics*, vol. 8, no. 1, pp. 25–54, 2021.

[14] T. Hiraoka, S. Takase, K. Uchiumi, A. Keyaki, and N. Okazaki, "Recurrent neural hidden Markov model for high-order transition," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 2, pp. 1–15, 2021.

[15] Z. S. Mirzazadeh, J. B. Hassan, and A. Mansoori, "Assignment model with multi-objective linear programming for allocating choice ranking using recurrent neural network," *RAIRO-Operations Research*, vol. 55, no. 5, pp. 3107–3119, 2021.

[16] M. Kowsher, A. Tahabilder, M. Z. Sanjid et al., "LSTM-ANN & BiLSTM-ANN: hybrid deep learning models for enhanced classification accuracy," *Procedia Computer Science*, vol. 193, pp. 131–140, 2021.

[17] H. Amich, M. Ben Mohamed, and M. Zrigui, "Multi-level improvement for a transcription generated by automatic speech recognition system for Arabic," *Annals of the American Thoracic Society*, vol. 16, no. 3, pp. 460–466, 2019.

[18] Y. Atmani, S. Rechak, A. Mesloub, and L. Hemmouche, "Enhancement in bearing fault classification parameters using Gaussian mixture models and Mel frequency cepstral coefficients features," *Archives of Acoustics*, vol. 45, no. 2, pp. 283–295, 2020.

[19] K. Wang, X. Liu, C. M. Chen, S. Kumari, M. Shojafar, and M. S. Hossain, "Voice-transfer attacking on industrial voice control systems in 5G-aided IIoT domain," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 7085–7092, 2021.

[20] X. Wu and Q. Lu, "Financial asset yield series forecasting based on risk-neutral fuzzy bilinear regression and probabilistic neural network," *Journal of Intelligent Fuzzy Systems*, vol. 40, no. 6, pp. 11829–11844, 2021.