Design and Application of "Vocational Skills+Mandarin" Remote Training Service Platform Based on Web Technology

Tao Zhang*, Zhenhong Li

Shandong Institute of Commerce and Technology, Jinan, Shandong, China

*550282973@qq.com

Abstract. With the in-depth development of language and writing work in the new era, the reform of Mandarin teaching has become a key link in the development of higher education, and the proposal of "vocational skills+Mandarin" plan has pointed out the direction for the development of college students' professional Chinese ability. In this regard, based on the current difficulties in the practice of Mandarin teaching in colleges and universities, this paper puts forward a design scheme of remote training service platform to promote the digital and intelligent transformation of Mandarin teaching system. Based on Web technology, the platform completes the overall structural framework, and integrates automatic speech recognition technology (ASR) and support vector regression algorithm (SVR) into it, realizing the platform's high integration of remote login, online training, practice evaluation, data analysis and other functional services. The test proves that the system can meet the actual needs of users at present, complete the deployment of Mandarin education and training process with digital information application as the core, and provide necessary technical support for promoting the application of "Vocational Skills+Mandarin" program.

Keywords: Web technology; Vocational skills+Mandarin; Education and training; ASR; Computer software application

1 INTRODUCTION

In the new period of building Socialism with Chinese characteristics, as an important development strategy of the country, "strengthening the promotion of national common language" is being given a brand-new development significance. From a macro perspective, it is related to social and economic development, historical and cultural heritage, and even national unity and national unity. For individuals, it directly affects our study, life, work and future development. [1] As an important position to popularize and standardize the use of Mandarin, colleges and universities shoulder the heavy responsibility of improving the ability of contemporary college students in Mandarin. However, at present, the traditional teaching mode is still used in Mandarin education

G. Guan et al. (eds.), *Proceedings of the 2023 3rd International Conference on Education, Information Management and Service Science (EIMSS 2023)*, Atlantis Highlights in Computer Sciences 16,
https://doi.org/10.2991/978-94-6463-264-4_10

in colleges and universities, with insufficient overall attention, monotonous and outdated teaching content, lack of professional teachers, lack of necessary practical courses and scientific assessment and evaluation mechanism, so that Mandarin education and teaching is closer to temporary training for examination, which greatly impacts students' learning enthusiasm and restricts the teaching effectiveness of the course. [2] In view of this, this paper holds that colleges and universities should fully realize their leading role in the popularization and standardized use of Mandarin, and earnestly practice the "vocational skills+Mandarin" plan to actively promote the implementation of Mandarin teaching reform in colleges and universities. Especially in the highly developed environment of educational informatization, it is necessary to build a Web-based vocational skills+Mandarin distance training service platform with the help of the practical advantages of digital information technology, so as to set up a new paradigm for the development of Mandarin education in colleges and universities and create a new model for the development of educational informatization in colleges and universities.

2 Overview of key technologies

2.1 Web technology

As a distributed network service based on the Internet, Web can support many users to query and browse information through a graphical and easy-to-access interface. [3] Web is a typical distributed application structure, and its application core lies in the transmission and sharing of data and information. Web technology refers to the general name of Web application construction technology, including Web server technology and Web client technology. Among them, Web client design technology mainly includes HTML language, Java Applets, script program, CSS, plug-in technology and VRML technology. Web server technology mainly includes server, CGI, PHP, ASP.NET and JSP technology. Through the combination of client-side technology and server-side technology, and with various application architectures of Web server, the division, decoupling and splitting of user requirements are realized, which makes the function of Web more and more powerful and can be applied to the design and development of most application systems.

2.2 Speech recognition technology

Speech recognition technology, namely Automatic Speech Recognition (ASR). As the most natural way of man-machine information interaction, the basic idea of speech recognition technology is to take speech as the research object, recognize and understand the input speech signal by the machine, compare the pronunciation quality models according to the pronunciation and vocabulary in the expert knowledge base, and complete the functions of scoring and error correction. The general workflow of speech recognition technology is: speech signal preprocessing, eigenvalue extraction,

language decoding and comparison, and recognition result output, as shown in Figure 1.

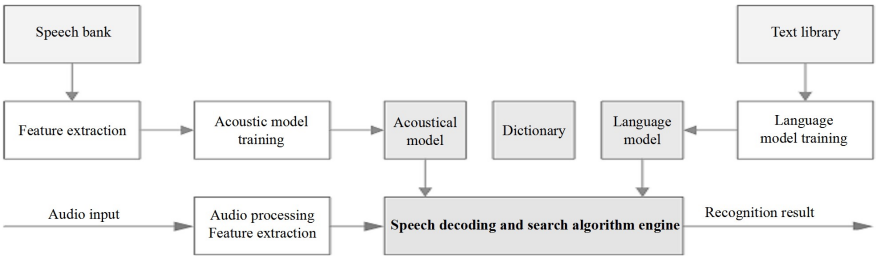

Fig. 1. The overall structure of speech recognition technology

A. Speech signal preprocessing

Speech signal preprocessing technology is an important premise and foundation of speech recognition technology, including signal digitization, framing, windowing, pre-emphasis and other steps. [4]

B. Mel-Frequency cepstral coefficients acoustic features

Mel-frequency cepstral coefficients (MFCC) is a way to describe the characteristics of speech signals. MFCC feature parameters use Mel frequency, which is obtained by transforming the linear frequency of speech signal. As shown in formula 1, it is the conversion relationship from linear frequency f to Mel frequency m.

$$m = 2595\lg\left(1+\frac{f}{700}\right) \tag{1}$$

The preprocessed speech signal is transformed by fast Fourier transform frame by frame, and the spectrum energy distribution is defined, and then it is transformed into Mel frequency domain by Mel filter. After logarithm, discrete cosine transform and differential extraction, the MFCC feature vector of each frame of speech signal is obtained. [5]

C. Acoustic model

Hidden Markov Model (HMM) is a statistical model in which the acoustic input of speech signal is associated with the basic unit of speech. [6] In order to improve the fitting distribution of state number and observation sequence probability in the parameter structure of HMM model, Gaussian mixture model is usually selected for estimation, that is, Gaussian mixture-hidden Markov model is formed, abbreviated as GMM-HMM model. The overall training method of GMM-HMM statistical framework is simple, changeable and adaptable, which can improve the convenience of ASR system application.

D. Search algorithm

After the acoustic model and the language model are constructed, a search algorithm network will be built by combining with the dictionary, and the input speech information will be aligned with the text content, and the Viterbi beam search algorithm will be used to complete the traversal search in the optimal path according to the WFST decoding mechanism, so as to complete the final speech content recognition. In WFST decoding mechanism, G stands for language model, L stands for dictionary, C stands for trisyllabic model, and H stands for acoustic model state. [7]

2.3 Development process

First of all, the platform belongs to B/S architecture, and the basic framework includes three parts: client, server and database. The client is a user interactive interface, which aims to facilitate users to call various data information and application services of the platform through graphical basic operations. Client page is built with JSP technology as the core, combined with HTML code, XHTML code, XML elements and JSP operation instructions, and integrated with Ajax technology to speed up the communication efficiency between client and server. Server-side development is more realized with the help of "request/response" Web development framework, and follows MVC design pattern, which connects the client-side interactive interface with the server-side under a specific data interface. [8] According to the actual requirements of this system, we choose SringMVC 4.1 development framework in Java language environment and Apache Tomcat 9.0 to complete the server-side configuration. In addition, the bottom operating system of the platform is Windows Swever Standard, the JDK version is 1.8.0_251, the integrated development tool is Eclipse Neon 4.6.2, and the database server is MySQL 5.7.

Secondly, remote login, online training, practice evaluation, data analysis and other functional services in the platform will be realized one by one based on the SringMVC framework on the server side of the platform. Among them, the practical evaluation needs to combine ASR technology to form a method class that can be directly called by the server, and fit the speech signal preprocessing, feature value extraction, language decoding comparison, recognition result output and other links. In the actual development process, you can directly use the new generation of Kaldi framework to quickly build a server-side ASR system. The advantage of Kaldi framework is that it can simplify complex speech signal preprocessing, speech database, dictionary processing, acoustic model training and language model training, and accelerate the realization of system function application. Taking speech signal preprocessing as an example, Kaldi framework can directly replace the traditional FFmpeg tool to complete audio file processing, directly determine the pre-emphasis coefficient of speech signal as 0.97 in script language, smooth the speech signal with a Hamming window with a length of 20ms, and complete the MFCC feature extraction of speech signal. In addition, the practice evaluation function will also be integrated into some evaluation models to evaluate and score the pronunciation quality of student users, that is, after the ASR system completes the speech recognition, it will continue to analyze and calculate the recognition results, obtain various evaluation characteristics needed for

the pronunciation quality evaluation, and finally output the pronunciation quality evaluation results in combination with the evaluation strategy. The pronunciation quality evaluation model includes GOP algorithm for pronunciation accuracy, pause length algorithm for pronunciation fluency, DTW algorithm for pronunciation intonation and comprehensive SVR pronunciation quality scoring algorithm. [9]

Finally, after all the platform functional applications are developed, they will be packaged and distributed to the server. After configuring the corresponding ports, users can complete the use of the system from the browser. Through the introduction of the above key technical theories, the overall environment of system development, the running process of related software and tools are determined, and the technical feasibility of the overall project of vocational skills+Mandarin remote training service platform is also clarified.

3 Functional implementation

3.1 Student side

A. Online training

Compared with the traditional offline training mode, the platform can make up for the shortage of professional teachers by online training. Under this module, the platform will upload video courses, ppt courseware materials, micro-courses, simulation tests and other forms of training content in advance according to the actual curriculum arrangement, so as to facilitate students to complete online learning. In addition, the digital teaching resources in the platform will be re-produced according to the teaching goal of "vocational skills+Mandarin", which will transform the original temporary and exam-oriented training into a long-term and systematic teaching practice, which can not only meet the actual needs of student users in work, study and socialization, but also promote personal sustainable development.

B. Practice evaluation

Under the plan of "Vocational Skills+Mandarin", student users should not only learn with online training courses, but also strengthen practical training to apply what they have learned. The platform supports student users to complete simulation training online, and with the help of speech recognition and evaluation model, the evaluation is automatically completed, further improving the efficiency of users' Mandarin learning.

Under this function module, student users can choose different practical training items according to their own habits or learning progress. The project includes basic training such as initials, finals and tones, as well as advanced training such as vocabulary, short reading and proposition expression. When the student user chooses to enter the practical training project, the platform will start the microphone of the user equipment to collect voice signals. Take reading a short passage as an example, there will be prompt text content on the platform interface, and student users can read it aloud. After the students finish, the system will automatically complete the speech

recognition and display the evaluation results of the students' Mandarin pronunciation quality.

The evaluation and scoring process of the platform for students' Mandarin pronunciation quality is to continue the phoneme segment recognition results from the Viterbi beam search algorithm, and extract various evaluation features from three aspects of Mandarin pronunciation standard, fluency and intonation. Finally, the vector regression algorithm is used to fuse the evaluation features of different dimensions, and the comprehensive score of students' reading practice training is obtained.

For pronunciation accuracy evaluation, it mainly depends on GOP algorithm. The calculation formula is shown in formula 2, where qi is phoneme, oi is observation vector, M is the total number of phonemes in GMM-HMM acoustic model, and P is likelihood. Similarly, the GOP score of a single phoneme can be normalized in the duration dimension, and the GOP score of the whole sentence can be obtained. [10]

$$GOP = \log \frac{P(o_i, q_i)}{\max_{j=1}^{M} P(o_i, q_i)} \quad GOP' = \frac{1}{N} \sum_{i=1}^{N} \frac{GOP(q_i)}{t_i} \tag{2}$$

In the face of fluency evaluation, it can be judged from the standards of reading speed, average stream length and average pause length of student users. As shown in Formula 3, the average pause duration is calculated, where T represents the total pause duration and N represents the total pause times.

$$L = \frac{T}{N} \tag{3}$$

The intonation evaluation needs to introduce pitch features, align the pitch features of the students' Mandarin pronunciation with the standard pitch sequence, and calculate the similarity between them, so as to judge the ups and downs of Mandarin pronunciation. [11] As shown in Formula 4, it is the calculation formula of DTW algorithm, where S represents the DTW distance between two groups of pitch sequences, Smin represents the minimum distance and the maximum similarity between them, and the final intonation evaluation score is W.

$$S_{\min}(i,j) = \min\{S_{\min}(i,j\text{-}1), S_{\min}(i\text{-}1,j), S_{\min}(i\text{-}1,j\text{-}1)\} + M(i,j), \quad W = \frac{100}{1 + a(S_{\min})^b} \tag{4}$$

Finally, the platform integrates three evaluation scores of accuracy, fluency and intonation, and SVR algorithm is used for effective fitting. As shown in Formula 5, it is the regression function of SVR. Among them, K is the kernel function, and the three feature scores are normalized by cubic polynomial function, and then enter the SVR model as input. After training, the values of parameters a and b are determined, and the feature fusion is completed, so as to achieve a comprehensive score on the pronunciation quality of students' Mandarin reading aloud.

$$f(x) = \sum_{i=1}^{m} (\alpha_i - \alpha_i^{\Phi}) K(x_i, x) + b \tag{5}$$

In order to verify the actual effectiveness of the platform's practice evaluation function, 2000 student users' voice information data were selected for simulation test, of which 70% were used as training sets and 30% as test sets. First of all, in the test of speech recognition accuracy, under the same acoustic characteristics of MFCC, the comparison results of HMM, GMM-HMM and end-to-end acoustic models in word recognition error rate are shown in Table 1. The results show that GMM-HMM acoustic model has the lowest error rate and the best recognition effect. Secondly, the evaluation results of the pronunciation quality of Mandarin reading are shown in Figure 2. The results show that the correlation between automatic platform scoring and manual actual scoring represents the effectiveness of platform functions. The higher the correlation, the stronger the platform evaluation performance.

Table 1. Identification results of the different acoustic models

Acoustical model	Misword rate
HMM	19.73%
GMM-HMM	16.72%
End-To-End	37.40%

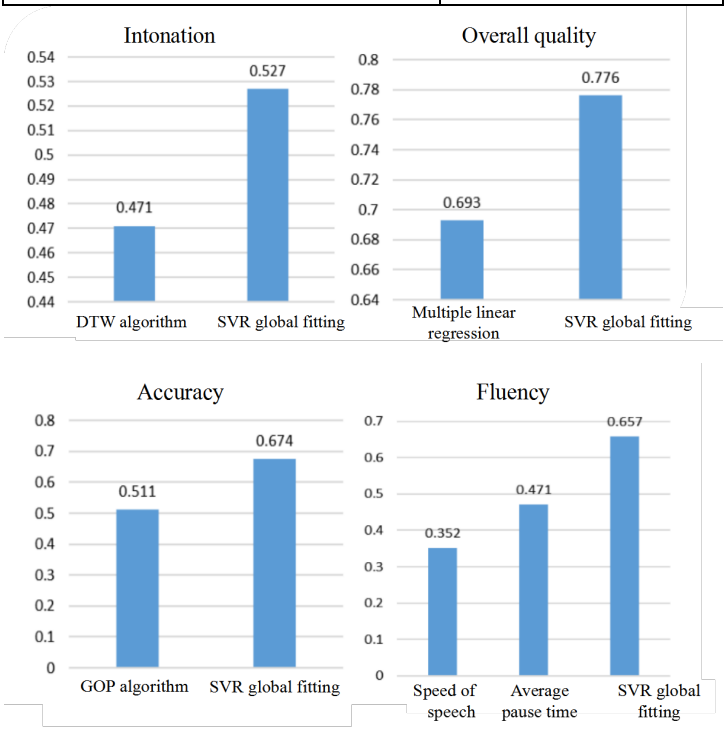

Fig. 2. Performance correlation test results of platform practice evaluation

3.2 Teacher side

Under the online teaching platform in colleges and universities, the functional authority of teacher users is mainly composed of three parts: student management, resource management and teaching management. As an important part of teaching management, assessment is an important criterion for judging the effect of students' Mandarin training. The platform can capture the complex learning behavior data generated by student users in the platform in real time, and construct the evaluation system standard of training effect, as shown in Table 2.

Table 2. Remote training effect evaluation system

Primary index	Secondary indicators	Evaluation standards
Learning attitude C_1	Study duration C_{11}	≧3 hours / time: 5 points ≧1 hour / time: 3 points ≦1 hour / time: 1 point
	Login frequency C_{12}	≧7 times / week: 5 points ≧5 times / week: 3 points ≦5 times / week: 1 point
Learning process C_2	Practice completion rate C_{21}	≧80%: 5 points ≦80%: 3 points ≦60%: 1 points
	Resource learning rate C_{22}	≧90%: 5 points ≦90%: 3 points ≦70%: 1 points
Learning ability C_3	Exam results C_{31}	≧90: 5 points ≦90: 3 points ≦60: 1 points
	Practice results C_{32}	≧85: 5 points ≦85: 3 points ≦60: 1 points

The platform compares each index value in pairs to determine its importance, and completes the construction of judgment matrix according to the provisions of comparative quantized values, as shown in Formula 6. According to the judgment matrix, the elements of each row are normalized by the AHP hierarchical analysis method and then summed, and the obtained row vector is normalized twice to get the ranking weight vector W, and the corresponding weight λmax is calculated by the sum-product method, as shown in Formula 7. [12] After the weight of each index value is determined, the platform automatically calculates the training effect score, and the simulation test results are shown in Table 3. The results show that the platform can conveniently and quickly complete the evaluation of teaching effect, correct the one-sidedness of traditional assessment methods, which provides a basis for improving the effectiveness of vocational education teaching.

$$C = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \tag{6}$$

$$\lambda_{\max} = \sum_{i=1}^{n} \frac{(CW)_i}{nW_i} \tag{7}$$

Table 3. Evaluation results of online learning effect

	Primary index	Secondary indicators	Weighted value	Item score	Average score
Learning effect score	Learning attitude C_1	Study duration	C_{11}=0.057	3.71	79.16
		Login frequency	C_{12}=0.105	3.63	
	Learning process C_2	Practice completion rate	C_{21}=0.076	4.00	
		Resource learning rate	C_{22}=0.079	3.87	
	Learning ability C_3	Exam results	C_{31}=0.166	87.1	
		Practice results	C_{32}=0.098	22.6	
	...	...	...		

4 Conclusions

In order to promote the reform of Mandarin teaching mode in vocational colleges, this paper puts forward a design scheme of Web-based distance training service platform based on many shortcomings faced by traditional teaching mode. The platform focuses on creating a new ecology of Mandarin education in colleges and universities from online training, content resources, practice evaluation and data analysis. In the follow-up research, the platform should further improve the accuracy of ASR system in practice evaluation, strengthen the realism of simulation practice, and make an attempt for the modernization and intelligent construction of higher education.

Acknowledgments

Subject: The Construction and Practice of Employment-oriented Vocational Skills and Mandarin Talent Training Mechanism in Higher Vocational Colleges YYWZTGJD2022C16

References

1. Zhou Wenwei. Analysis of the Problems in the Popularization of Mandarin in the New Period[J]. Shenhua.07 (2020).
2. Yang Youmian. Exploration of Mandarin Teaching in Colleges and Universities Based on Mandarin Proficiency Test[J]. Ability and Wisdom.06 (2022).
3. Jia Lina. Research and Application of Performance Optimization of Analytical Web Front-end[J]. Electronic Test.08 (2022).
4. Sun Lei. Research on Preprocessing Method of Multi-person Dialogue Speech Recognition in Complex Acoustic Scene[D]. University of Science and Technology of China.05 (2020).
5. Zhao Yangqing, Peng Zhicai, etc. Feature Extraction Process of Mel-frequency Cepstral Coefficients in Audio. [J]. Information Technology and Informatization.01 (2023).

6. Fang Min. Speech Recognition Technology Based on HMM[J]. SP.03 (2018).
7. Wang Kai. Research on Speech Recognition Based on Kaldi[D]. Nanjing University of Posts and Telecommunications.11 (2021).
8. Qi Xuedong, Han Lifeng. Design and Application of Web Framework Based on MVC[J]. Microcomputer Applications.03 (2021).
9. Zhang Long. Research on Automatic Evaluation Method of Pronunciation Quality of Mandarin Chinese[D]. Harbin Institute of Technology.09 (2014).
10. Xu Lili. Research on Speech Recognition and Speech Evaluation Technology in Mandarin Proficiency Test System[J]. Course Education Research.05 (2017).
11. Shen Caifeng. Research on Tone Evaluation Algorithm of Mandarin Chinese[D]. Soochow University.10 (2012).
12. Cheng Zhenlin. Analysis of Network Learning Behavior Supported by Multimodal Data[J]. China Adult Education.04 (2022).