Overcoming the Difficulty of Teaching Chinese Pronunciation Based on Artificial Intelligence Models

Xinyu Lin
Computer Science and Technology, Fujian University of Technology, Fujian, China

Keywords: Speech Recognition, Pronunciation Correction, Artificial Intelligence.

Abstract: In recent years, the application of Artificial Intelligence (AI) in language teaching has attracted increasing attention, especially in pronunciation. This is because language plays a fundamental role in communication, and oral communication is particularly important in daily life. However, for students learning Chinese abroad, it is difficult to create a good learning environment due to the limited opportunities to contact Chinese people and they cannot learn in their mother tongue. Therefore, pronunciation is also one of the biggest problems. Artificial intelligence has brought new opportunities and educational methods to the educational community, and its application is increasingly widely used in the field of language teaching. This paper focuses on the development of AI in Chinese pronunciation teaching. This paper first discusses the principles of AI and its application in various fields including medical biology, sound and identification, and agricultural environmental protection. Then this paper focus on the application of artificial intelligence in Chinese pronunciation teaching, especially in the detection and correction of speech errors. The paper discusses the use of automatic speech recognition techniques to judge the correctness of pronunciation, and the ways to correct these errors. This paper also highlights the challenges that AI faces in detecting and correcting Chinese pronunciation errors, including the lack of emotional output in AI speech synthesis. Finally, this paper concludes that researchers should continue to optimize algorithms and models to improve accuracy and explore AI-assisted language teaching for different languages.

1 INTRODUCTION

In today's world, an increasing number of schools not only pay attention to grade tests, but also pay more attention to oral communication in foreign language teaching. Language serves as a fundamental bridge for communication, with oral interaction playing a predominant role in daily life. While Chinese plays an important role in the world. China has a rich cultural heritage, and its excellent traditional Chinese culture has also attracted many foreign friends. Therefore, mastering Chinese well is very beneficial to learning Chinese culture.

However, for students studying Chinese abroad, there are fewer opportunities to contact Chinese people, and they do not study in their mother tongue, which makes it challenging to create a good learning environment. Therefore, pronunciation has become one of the biggest problems. With the progress of science and technology, the application of artificial intelligence is more and more widely used in the world. Artificial intelligence has brought people a convenient way of life, brought a new development direction to society and the country, and also brought opportunities and new ways of education to the educational community.

Artificial intelligence is widely used in many domains such as medical biology, sound and identity identification, and agricultural environmental protection. It can analyze complicated medical data that can be used for diagnosis, treatment and predict outcome in many clinical situations, and can also improve the performance of hand inherent and demographic features based on ANFIS-SC algorithm and SVM-ECOC algorithm (Ramesh et al. 2004. Abdullahi et al. 2022). In recent years, new progress has also been made in the combination of artificial intelligence and language learning. Guo et al. adopts intelligent speech technology to assist Tibetan students to learn Mandarin, and reduces the error rate and improves the accuracy rate through three steps of speech recognition, speech synthesis and direct algorithm (Guo et al. 2019). In addition, because there are many dialects in different regions and they have

Lin, X.
Overcoming the Difficulty of Teaching Chinese Pronunciation Based on Artificial Intelligence Models.
DOI: 10.5220/0012837600004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 334-338
ISBN: 978-989-758-690-3

unique speech characteristics, combining Hidden Markov Model (HMM) and Viterbi algorithm can also cut the speech, which is helpful to improve the construction efficiency of mandarin dialect speech database (Lai 2022). Although AI has made considerable progress in the language field, its application in speech teaching is still relatively rare. Therefore, it is necessary to make a relevant summary of the current progress of artificial intelligence in the field of language teaching.

The rest of this article is organized as follows: First, in the second part, this review will explain the principles of AI; then, in the third part, the application of AI in Chinese pronunciation teaching will be discussed; finally, this paper summarizes the whole paper and propose the direction for future improvement.

2 METHOD

2.1 The Framework of Artificial Intelligence

Artificial intelligence has many algorithms, among which machine learning is the most widely used. The core idea of machine learning is to simulate the human learning process, through continuous training and optimization, enabling the machine to automatically identify, classify, predict and reason. The specific process includes three main stages, namely data preparation, model selection and training, and model evaluation and optimization. A sample of the machine learning workflow in real applications is shown Fig. 1.

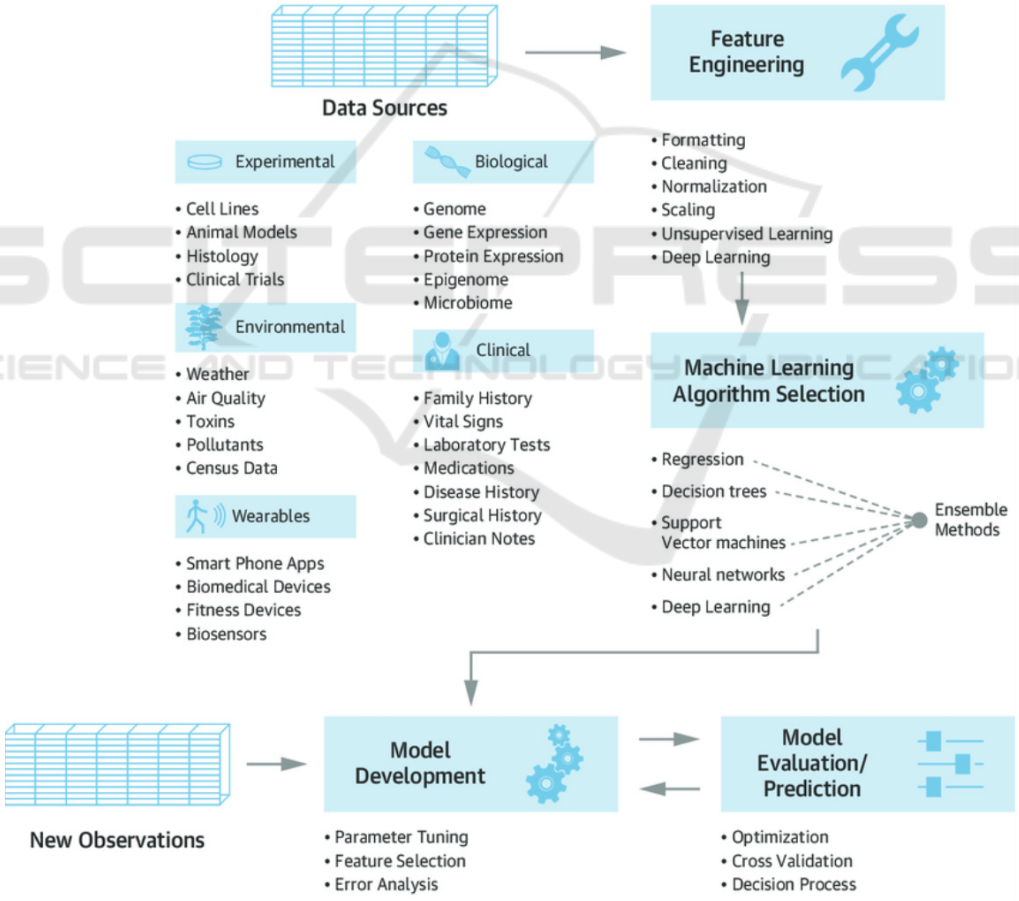

Figure 1. The workflow of machine learning in the real applications (Johnson et al. 2018).

First, in the data preparation stage, it is required to collect and clean the data, while conducting feature selection and extraction to better represent the problems and patterns. Next, in the model selection and training stage, the machine learning model suitable for the problem should be chose. After selecting the model, the data is divided into training and test sets, and used the training set to fit the model.

Model fitting was performed by tuning the parameters of the model so that it can best fit the training data. Then, the model evaluation and optimization are conducted to evaluate the performance of the model and tune it. the model performance can be evaluated using various indicators. If the model does not perform, it is possible to adjust the hyperparameters of the model, add more data or try other models. Finally, after the model evaluation and optimization, the trained model can be applied to the new data for prediction and inference.

2.2 AI-Based Pronunciation Detection

Some researchers use automated speech recognition (ASR) technology shown in Fig. 2 to build a system (Kholis 2021). It has proved more useful in many areas, such as improved speech comprehension, speech therapy, and pronunciation perception training (Badin et al. 2010, Fagel & Madany 2008, Rathinayelu et al. 2007). Now it is possible to use the automatic speech recognition technology in artificial intelligence to judge whether the pronunciation of the second language learners is correct in various systems, such as Saybot , spelling system, etc (Chevalier 2007, Morton et al. 2010).

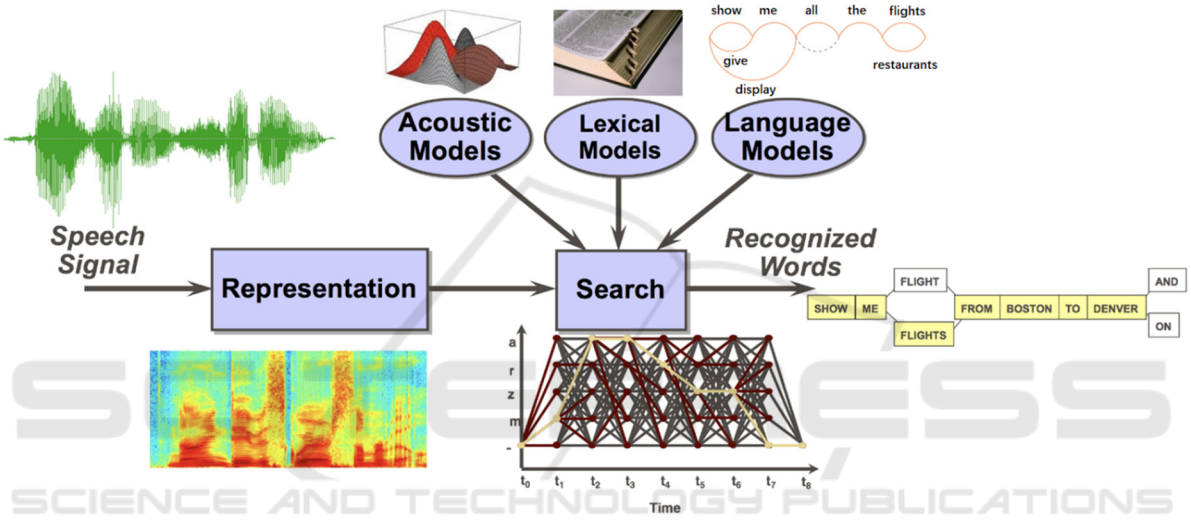

Figure 2. The schematic diagram of automated speech recognition (MIT CSAIL).

In recent years, the application of deep neural networks (DNNs) in speech recognition is also very common. The architecture of neural network is shown in Fig. 3. Compared with the traditional speech processing technology, it can better improve the accuracy and efficiency of detection, and has a good prospect in improving the empirical performance (Hu et al. 2013). Li et al. proposed the use of phonetic attributes to solve the detection of pronunciation errors and provide diagnostic feedback. They first measure speech quality at the subsegmentation level using speech attribute scores, and then integrate them into neural network classifiers to generate segmentation articulation scores (Kun et al. 2016). Guo et al. Proposed a Chinese-Tibetan interlanguage speech synthesis method based on deep neural network (DNN), using speaker adaptation training. The initial model and the final model were used as speech synthesis units in Mandarin and Tibetan to train a set of DNN-based average speech models (AVM) (Guo et al. 2018). However, so far, the latest research based on DNN mainly focuses on the mainstream English, Japanese and other second language learning, and lacks the learning of Chinese pronunciation. Therefore, there are few studies on the detection and correction of Mandarin pronunciation errors integrating DNN.

2.3 AI-based Pronunciation Correction

After detecting the pronunciation errors, the pronunciation errors should be corrected. Guo et al. uses speech synthesis to synthesize the correct pronunciation of incorrect speech (Guo et al. 2019). The most widely used methods include waveform-connected, statistical parameter speech synthesis based on HMM (Clark et al. 2007, Zen et al. 2009). At the same time, the oral rate is adjusted to the straight algorithm to help better understand Mandarin and correct pronunciation. The specific implementation is to use the speaker to adapt to the method training.

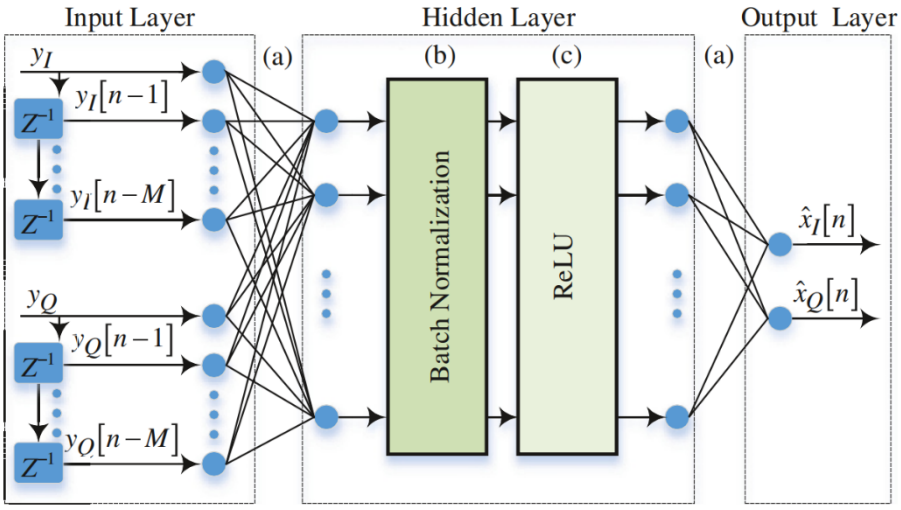

Figure 3. The architecture of neural network (Jalili et al. 2022).

3 DISCUSSION

In the field of artificial intelligence application in Chinese teaching, although many researchers have explored it, the related products on the market are still relatively scarce. This is mainly due to two reasons: first, the global circulation language is still mainly English, leading to the dominance of English products in the market; second, the target audience of Chinese pronunciation teaching products is relatively small, mainly including children in rural areas of China and foreign friends who love Chinese culture. However, for these target audiences, the high price of AI products has become a big problem, especially for rural children, who often face payment difficulties.

At the same time, AI also faces challenges in detecting Chinese pronunciation. When people use artificial intelligence, users cannot understand how to make decisions behind artificial intelligence, and users cannot understand this process, which indirectly leads to users cannot really understand the deep logic, which will greatly improve the difficulty of understanding. At present, the detection accuracy of AI is not 100 percent, and sometimes there is a misjudgment. This involves the problem of key algorithms, and how to improve the relevant models and algorithms to improve the detection accuracy is an important subject to be studied. Some advanced models or algorithms could be considered in this case (Qiu et al. 2022, Chau 2024). At present, there are relatively few research on algorithms and models in this area.

In addition, in the Chinese pronunciation correction link, how to improve the accuracy of the correction has become a key problem. Constantly optimizing AI algorithms to improve correction accuracy is the core task. Through continuous technological innovation and in-depth research, it can be expected to overcome these challenges and further promote the application and development of AI in Chinese teaching. Finally, the AI can detect the wrong pronunciation and correct it, resulting in correct phonetic guidance. However, because the speech output of AI is mechanical and lacks emotion, the guiding effect may not be obvious for the user. In order to improve the user experience and teaching effect, it is possible to input human voice to synthesize voice and output to the user.

4 CONCLUSION

In this paper, a review of AI in teaching Chinese pronunciation was provided. First, the principle of artificial intelligence, then the application of artificial intelligence in the detection and correction of Chinese pronunciation. For example, the use of speech properties to solve the detection of speech errors, provide diagnostic feedback, and use in speech synthesis to synthesize the correct pronunciation of incorrect speech. This article focuses on summarizing the relevant research made by researchers in the field of Chinese pronunciation teaching, and also has a good summary of each work, which can play a good reference role for subsequent readers. However, the article currently only focuses on the teaching of Chinese pronunciation and does not summarize and explain the teaching of other languages, and some

traditional AI-assisted teaching methods are not mentioned. In the future, AI should constantly optimize algorithms and models to improve accuracy. At the same time, researchers should study the application of artificial intelligence in various language teaching in the world, improve the ability of artificial intelligence to assist language teaching, and promote people from all over the world to learn different languages, so as to better promote cultural communication.

REFERENCES

A.N. Ramesh, C. Kambhampati, J.R.T. Monson, et al., Ann. R. Coll. Surg. Engl. **86**(5), 334 (2004)

S.B. Abdullahi, C. Khunpanuk, Z.A. Bature, et al., IEEE Access **10**, 49167-49183 (2022)

W. Guo, H. Yang, Z. Gan, *Improving Mandarin Chinese Learning in Tibetan Second-Language Learning by Artificial Intelligent Speech Technology*, Proc. Int. Joint Conf. Inf. Media Eng. (IJCIME) (2019)

Y. Lai, Mobile Inf. Syst. (2022)

K.W. Johnson, J. Torres Soto, B.S. Glicksberg, K. Shameer, R. Miotto, M. Ali, E. Ashley, J.T. Dudley, J. Am. Coll. Cardiol. **71**(23), 2668-2679 (2018)

A. Kholis, Pedagogy: J. Engl. Lang. Teach. **9**(1), 01-14 (2021)

P. Badin, Y. Tarabalka, F. Elisei, G. Bailly, Speech Commun. **52**, 493-503 (2010)

S. Fagel, K. Madany, *A 3D virtual head as a tool for speech therapy for children*, Proc. Interspeech (2008)

A. Rathinavelu, H. Thiagarajan, A. Rajkumar, *Threedimensional articulator model for speech acquisition by children with hearing loss*, Proc. 4th Int. Conf. Univ. Access Hum.-Comput. Interact. **4554**, 786-794 (2007)

MIT CSAIL, Automatic Speech Recognition https://www.csail.mit.edu/research/automatic-speech-recognition

S. Chevalier, *Speech interaction with Saybot: a CALL software to help Chinese learners of English*, Proceedings of the SLaTE-2007 workshop, 37–40 (2007)

H. Morton, M.A. Jack, Comput. Assist. Lang. Learn. **4**(23), 295–319 (2010)

W. Hu, Y. Qian, F.K. Soong, *A new DNN-based highquality pronunciation evaluation for computer-aided language learning (CALL)*, In Proceedings of the Annual Conference of the International Speech Communication Association, 1886–1890, Lyon: International Speech Communication Association (2013)

L. Kun, X.J. Qian, H. Meng, IEEE/ACM Trans. Audio, Speech, Lang. Process. **25**(1), 193–207 (2016)

W.T. Guo, H.W. Yang, Z.Y. Gan, *A DNN-based Mandarin-Tibetan cross-lingual speech synthesis*, 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE (2018)

R.A.J. Clark, K. Richmond, S. King, Speech Commun. **49**(4), 317–330 (2007)

H. Zen, K. Tokuda, A.W. Black, Speech Commun. **51**(11), 1039–1064 (2009)

F. Jalili, Y. Zhang, M. Hintsala, O.K. Jensen, Q. Chen, M. Shen, G.F. Pedersen, IET Microwaves, Antennas & Propag. **16**(1), 62-77 (2022)

Y. Qiu, J. Wang, Z. Jin, H. Chen, M. Zhang, L. Guo, Biomed. Signal Process. Control **72**, 103323 (2022)

H.N. Chau, T.D. Bui, H.B. Nguyen, T.T. Duong, Q.C. Nguyen, IEEE/ACM Trans. Audio, Speech, Lang. Process (2024)