The Practice of College English Teaching Reform with Intelligent Voice Evaluation in the Context of Artificial Intelligence Empowering Education

Yamei Zhang[(✉)]

Department of General Courses Teaching, Shandong Institute of Commerce and Technology, Jinan, China
20010336@sict.edu.cn

Abstract. Artificial intelligence (AI) technology empowers the digital transformation of education, which is gradually shaping a new form of education. Based on the theory of "computer-assisted language learning", an overall teaching reform practice of human-machine collaborative education was conducted through the deep integration of artificial intelligence and college English teaching in the aspects such as teaching mode, teaching evaluation, and promoting teaching by competition. With the assist of the FIF oral language training system, the intelligent speech assessment was achieved in students' oral practice by the means of multidimensional speech evaluation, dependency syntax analysis and error detection coding analysis. In the test sample in this paper, 95% of the students' oral pronunciation errors were identified, the time spent by students in learning was increased by 15%, the evaluation score of oral English was increased by 5%, and the student satisfaction rate reached 99%. The results of this study show that the application of artificial intelligence technology in English teaching improves teaching effectiveness, enhances students' enthusiasm and initiative in learning, and provides feasible solutions for smart teaching in teaching, learning, evaluation and testing.

Keywords: The Artificial Intelligence · College English Teaching Reform · Intelligent Voice Evaluation

1 Introduction

The development of modern information technology has had a profound impact on education. Among them, new technologies with artificial intelligence and big data as the core are gradually changing the teaching mode.

The application of computers in English language teaching represents the beginning of the introduction of information technology into language teaching. Computer-assisted language learning (CALL), widely used in English language academia, was defined as "the study of the application of computers in language teaching and learning" [1]. The earliest application of "computer-assisted language learning" was traced back to
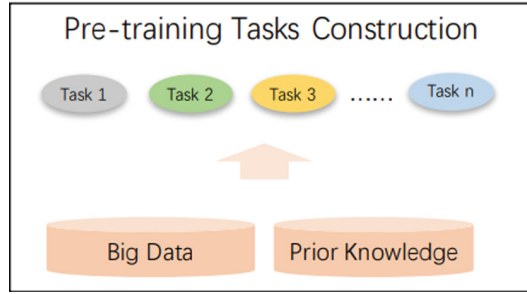
Fig. 1. Pre-training Tasks Construction

the 1960s. When the computer just appeared in the form of a giant mainframe in the university, the PLATO project of the University of Illinois in the United States began to develop "computer-assisted language teaching" system.

With the application of new information technology, natural language analysis based on artificial intelligence and big data has begun to attract academic attention. Academic circles continue to improve natural language processing methods to present parts of speech and semantic relationships through lexical analysis, dependency syntax analysis, semantic role tagging, semantic dependency (dependency tree) analysis, and semantic dependency (dependency graph) analysis. Various improvements and innovations proposed for problems in pre-training language models. ERNIE 2.0: A Continual Pre-Training Framework for Language Understanding [2]. This work introduces many tasks in the pretrain stage to assist the learning in the pretrain stage. Its core point lies in the proposed methods of constructing pre-training tasks, including Knowledge Masking Task, Capitalization Prediction Task, Token-Document Relation Prediction Task, etc. (Fig. 1).

China has incorporated educational informatization into the strategy of strengthening the country through education, and issued the ten-year development plan for educational informatization (2011–2020) and the educational informatization 2.0 action plan to support and lead the modernization of education with educational informatization [3]. Colleges and universities has developed disciplines and professional systems related to artificial intelligence in Colleges and universities, to explore the talent training mode of "artificial intelligence+x", and to strengthen the training of compound and applied talents.

In 2007, the Ministry of education pointed out in the "College English course teaching requirements" that the development of information technology with computer network as the core has provided technical support for teaching reform, and it was clearly proposed to reform the traditional teaching mode with advanced information technology [4]. Therefore, with the in-depth development of College English teaching reform, digital teaching based on computer and network has become the main practical trend of College English classroom teaching innovation [5]. The active introduction of new information technologies such as artificial intelligence has become one of the hot spots in College English teaching research.

This paper uses speech evaluation technology and error detection coding analysis to intelligently evaluate students' oral language training in terms of accuracy, fluency, integrity, tone accuracy and other evaluation dimensions. Combined with the data of students' learning process, it makes a useful exploration of artificial intelligence technology in human-computer collaborative teaching.

This study provides a multi-dimensional evaluation method for the application of artificial intelligence in oral English training, which is of great significance.

2   Teaching Reform of the Integration of Artificial Intelligence and College English Teaching

The College English should be actively integrated into the digital transformation, complied with the development trend of information technology, to make full use of AI technologies to promote the in-depth integration and development of information technology and classroom teaching.

2.1   Reform of Human-Computer Collaborative Education Teaching Mode

Based on emerging technologies such as artificial intelligence and big data, the mode reform and ecological reconstruction of education are supported by new technologies. Various forms of teaching resources, equipped with mobile Internet, artificial intelligence and other technical platforms, can realize an output in an intelligent, personalized, situational and accurate way. Simultaneously, students can realize personalized autonomous learning with the help of artificial intelligence technology.

The self-learning software was selected to put into teaching practice, such as YOUMOOC, AI translation, Iwrite network, Unipus APP, English dubbing APP, FIF oral training system. Combined with course content, the personalized design was carried out, the corresponding pronunciation training was formulated. The situational communication and expressing thinking were proposed to promote students' knowledge construction and innovative application.

The knowledge map established by artificial intelligence conducted multi-dimensional analysis of students' spoken and written content, and the automatic scoring and correction were offered immediately, so that students could know which aspects need to be strengthened, and effective personalized learning and guidance were provided based on big data.

The automatic evaluation and personalized tutoring functions completed by the machine provided a personality diagnosis, which corrected the usage of vocabulary, sentences and discourse, and offered the recommended expressions. By assisting teachers to provide personalized guidance to students, the more accurate and effective teaching practice was fulfilled.

2.2   Teaching Evaluation Reform

AI technology was used to improve the efficiency of English learning and evaluation, so that the multidimensional and three-dimensional evaluation was carried out.

Taking the FIF oral training system as an example, its speech evaluation was realized by a software system that automatically evaluates the pronunciation level, pronunciation errors, defect location and problem analysis through intelligent speech technology.

The technical principles of speech evaluation involve phoneme sequence, phoneme boundary segmentation, acoustic model calculation of phoneme posterior probability, threshold detection, etc.

Take the word "student" for instance, here was a phonetic evaluation on the word pronounced in the following way. The evaluation text was "s/tu/den/t". If a real pronunciation was "s/tu/den/d", according to the technical principle of voice evaluation, the phoneme sequence and the phoneme boundary segmentation were correct, but the phoneme posterior probability calculated by the acoustic model, was 0.91/0.94/0.95/0.10 respectively. The phoneme posterior probability value ranges from 0 to1. The higher the value, the more accurate the pronunciation. On the contrary, the lower value shows a greater the deviation from the correct pronunciation.

- For example: student;
- Evaluation text: s/tu/den/t;
- Real pronunciation: s/tu/den/d;
- Phoneme posterior probability: 0.91/0.94/0.95/0.10.

It was seen from the above test examples that the pronunciation probability value of the corresponding /t/ was low, down to 0.10, and this false pronunciation was detected in combination with the threshold method.

The evaluation dimensions of speech evaluation include the following aspects:

- Accuracy: pronunciation accuracy (average posterior probability of phonemes);
- Fluency: no catnap, smooth reading (speaking speed, increased reading ratio, improper pause ratio);
- Completeness: whether to read aloud completely (missing reading ratio);
- Tone accuracy: pronunciation accuracy of tones;
- Other dimensions: standard degree, continuous reading, stress, intonation, etc.

The artificial intelligence speech technology based on FIF oral training system realized scientific evaluation, quantitative analysis and real-time feedback.

3   AI+ Online Competition, Promoting Learning and Teaching

With the original intention of "promoting learning through competition and teaching through competition", a new mode of college English teaching under the empowerment of artificial intelligence technology was put into teaching practice. With the FIF oral English competition, the multidimensional model of speech detection and the analysis of speech error detection code were applied to English teaching.

3.1 Application of AI+ Online Oral Challenge

This competition adopted online competition, intelligent scoring and instant feedback, which greatly improved the enthusiasm and participation of students.

This competition had double evaluation of "AI intelligent scoring" and "expert manual scoring", and the final ranking was sorted from high to low according to the total score of the double evaluation.

From the statistics of competition, the cumulative number of challenge levels of the oral challenge was more than 5000, and the challenge duration was a total of 15000 min, which showed that the students' enthusiasm for learning and participating was increasing. Students actively completed oral tasks such as follow-up imitation, scenario simulation, role play, and the like. As a result, the students became more confident in oral expression.

3.2 Customization of Digital School-Based Oral Training Resources

With the help of "FIF oral training system", the online teaching mode for College English listening and speaking courses was established based on the rich school-based customized resources of the system. By arranging personalized learning tasks around pre-class, in-class and after-class, and AI intelligent scoring with real-time feedback, the system created a good English online learning environment and teaching atmosphere for teachers and students.

3.2.1 Students' Independent Learning

Oral training forms included word follow-up, sentence follow-up, paragraph follow-up, and role play.

As to students' Independent Learning, there were active students shown by the experience value from the three aspects: the length of study, the number of ability points, and the number of challenging exercises.

The big data recorded students' learning behavior as following:

- Complete the task: 236 people-time;
- Study independently: 1059 people-time;
- Post interactions: 124;
- Challenge levels: 3364;
- Cumulative recording: 176 h;
- Per capital: 26 min.

Student oral ability value trend and students' oral ability dimension were demonstrated with numbers from 0 to 5, of which the bigger the number is, the more accurate the pronunciation is. The average value was recorded as following:

- Pronunciation accuracy: 3.1;
- Reading fluency: 4.2;
- Vocabulary and grammar: 3.5;

- Flexible communication: 3.7;
- Discourse length and coherence: 3.4;
- Lexible communication: 3.2.

3.2.2 Online Digital Resources

According to the statistic, online teaching resources were established for improving students' oral ability, and presented to the students accurately. The chart below shows the establishment and use of online digital resources.

AI Education and AI teaching assistant played an important role in the competition. Speech synthesis supported teachers to build exercises for a total of 94 times. Speech evaluation supported students' oral training for a total of 10000 times. Online resources were recommended, and oral training tasks were assigned, of which they were:

- CET-4 and CET-6 oral English;
- Vocabulary training;
- Voice training;
- Workplace spoken English;
- Grammar training;
- TOEFL and IELTS.

3.3 Statistical Data of FIF Oral Training

3.3.1 Daily Data of Competition Process

It can be seen from Table 2 that the statistical data of FIF oral training were recorded from the following five aspects, involving Statistics Time, Number of participants, Number of challenge levels, Challenge duration (min) and Average challenge time (min).

In Table 1, the challenge duration (min), and the duration of a single day reaches 3998.58 min from 744.22 at the beginning, and the average challenge duration (min), from 12.83 to 26.31.

Table 1. Daily data of competition process

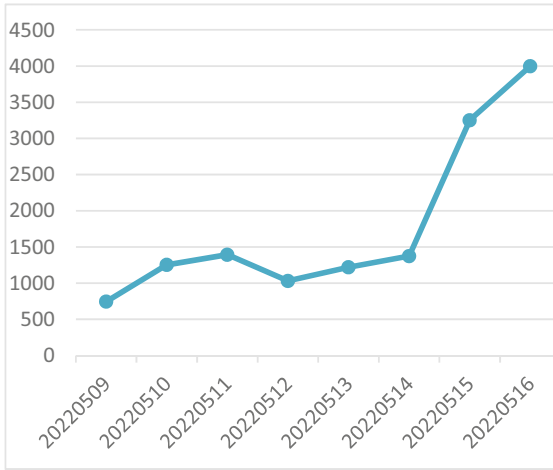| Date	Number of participants	Number of challenge levels	Challenge duration (min)	Average challenge time (min)
20220509	58	204	744.22	12.83
20220510	94	365	1252.1	13.32
20220511	91	430	1393.35	15.31
20220512	87	339	1030.75	11.85
20220513	84	386	1219.52	14.52
20220514	75	463	1374.57	18.33
20220515	143	967	3251.23	22.74
20220516	152	1252	3998.58	26.31

Fig. 2. Challenge duration

Table 2. Voice evaluation dimension

Number of participants	Accuracy average	Fluency average	Integrity average
58	61.96	75.54	88.75
94	70.57	83.86	95.83
91	68.96	82.33	94.96
87	66.4	78.67	89.05
84	72.8	85.47	96.37
75	74.64	87.26	96.5
143	63.89	79.52	94.52
152	62.35	78.22	93.77

3.3.2 FIF Speaking Training Statistics Graph

Figure 2 shows FIF speaking training statistics graph about challenge duration (min). The abscissa is the date, and the ordinate is a numerical value.

The blue line in Fig. 2 represents the challenge duration (min), and the single-day duration ranges from 744.22 on May 9 to 3998.58 min on May 16.

The data shows with the longer time of autonomous learning, evaluation and motivation received from training enabled students to maintain high learning initiative.
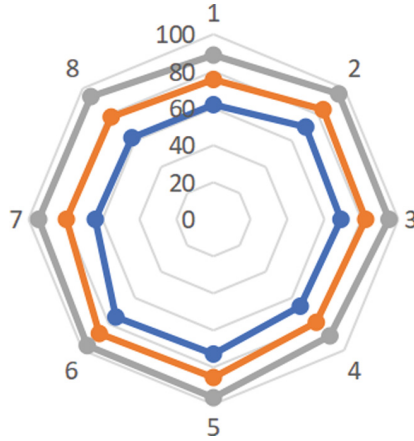
Fig. 3. Pronunciation dimension radar chart

3.3.3   Intelligent Voice Evaluation

The subjects' pronunciation was evaluated from the following three dimensions: pronunciation average, fluency average and integrity average (Fig. 3).

In the above pronunciation dimension radar chart, blue, orange and gray represent the average scores for accuracy, fluency and integrity, respectively.

As can be seen from the figures, among the three dimensions of pronunciation, integrity generally has the highest score, followed by fluency. While, accuracy has the lowest score. This showed that through training, students had the ability to read sentences completely and fluently, but there were generally many problems in pronunciation, and targeted reinforcement training was needed.

3.3.4   A Sample of Voice Evaluation

In actual language evaluation, the accuracy of pronunciation is intuitively displayed through different colors.

As is seen in the sample of Fig. 4, green represents excellent pronunciation, orange represents good pronunciation, red represents pronunciation error, and black represents unread.

In the example in Fig. 4, the total score 95 was for overall speech. There was an error detected and shown in red. "Olympic" was marked in red, representing pronunciation error.

Fig. 4. A sample of voice evaluation

3.3.5   Speech Error Detection Code

The coding analysis for the phonetic error detection is as follows:

Error (serr_msg)

serr_msg=0: normal [əˈlɪmpɪk]

serr_msg=1: syllable mispronunciation ([əˈlɪmpɪk] is pronounced as [əlɪmpek])

serr_msg=2048: unstressed (Premise: syll_accent=1)

serr_msg=2049: unstressed and syllable mispronounced (Premise: syll_accent=1)

The above example showed that the multi-dimensional evaluation of intelligent speech assessment, speech error detection coding analysis and the visual display were helpful for students to find problems and correct them in time when training, and also helpful for teachers to summarize common pronunciation problems for more effective teaching.

Artificial intelligence gave full play to the role of teaching tools, which greatly provided teaching and learning effectiveness. Compared with traditional teaching, the student satisfaction of teaching with AI technology reached 99%. Oral English evaluation scores increased by 5%; students' time devoted to learning increased by 15%. The quality of student learning and satisfaction both improved significantly.

4 Conclusions

Under the concept of "Artificial Intelligence Empowering Education and Teaching", the useful explorations were completed in the deep integration of artificial intelligence and college English teaching from the teaching model, the teaching evaluation and the teaching process and results.

Aiming at students' oral language training, the voice evaluation technology of the FIF oral language training system was conducted to intelligent real-time evaluation, multi-dimensional evaluation and speech error detection coding analysis. From the data of oral training process, it was obvious that the students put longer time into learning, and correspondingly, the high scores were achieved. It was shown that the quality of student learning and satisfaction have improved significantly.

To summarize, the English teaching mode of human-machine integration, which is empowered by artificial intelligence, has been proved to be effective in practice. The deep integration of big data technology and teaching process is an important technical direction of English teaching reform in the future.

References

1. Levy M (1997) Computer-assisted language learning: context and conceptualization. Oxford University Press, New York
2. Sun Y, Wang S, Li Y et al (2019) ERNIE 2.0: a continual pre-training framework for language understanding. https://www.researchgate.net/publication/334760209
3. Xi JP (2022) Continue to make my country's digital economy stronger, better and bigger. Qiushi J 2:1–3
4. Xu B (2018) Reflections on the construction of the training system for technical and technical personnel in higher vocational colleges. J Beijing Finan Trade Vocat Coll 03:66–68
5. Xu D (2021) Challenges and countermeasures for vocational education technical skilled talents training in the digital economy era. Res Industry and Technol Forum 20:109–110