

Study of Artificial Intelligence-Assisted English Oral Teaching

Jing Wang¹, Jieren Zhang², Lina Chen², and Jianbiao Dai^{1,*}

¹ Institute for Data Engineering and Science, University of Saint Joseph, Macao, China

² Shanghai Netclass IT Co., Ltd., Shanghai 200436, China

*Corresponding Author: dai.jian.biao@usj.edu.mo

Abstract

This study investigates and practices the auxiliary role of Artificial Intelligence (AI) technology, especially Large Language Models (LLM) and Text-to-Speech (TTS) technology, in English oral teaching. In the study, an AI-based LLM oral dialogue platform was constructed to replace teachers for one-on-one interactive training with students. It has been applied in college English teaching in various universities, and teaching data has been collected and analyzed in real-time. The effectiveness of AI technology in assisting college English oral teaching was evaluated, and the role of AI technology application in improving students' oral skills was explored. We conducted comparative analysis on indicators such as the speaking sentence length, speech rate, and combined with questionnaires from students, we surveyed the practical effects of AI oral teaching. The study shows that AI-assisted dialogue training has a significant positive effect on college English oral skills, and we have also proposed corresponding teaching strategies and suggestions.

Keywords

Artificial Intelligence; Large Language Models; Text-to-Speech; Sentence Length; Speech Rate.

1. Introduction

With the advancement of globalization and the increasing frequency of cross-cultural communication, the importance of English speaking ability is becoming more and more prominent. Traditional college English teaching models often focus on grammar, vocabulary, and reading comprehension, while neglecting the cultivation of speaking communication skills. The lack of dialogue training has always been a problem in speaking teaching. College English teaching classes are often large, and teachers cannot give enough attention to each student. As a result, many students may not have enough opportunities to practice speaking in class. This makes them seek other speaking practice opportunities outside the classroom, but such opportunities are often hard to find. In addition, many students may be reluctant to actively participate in class due to shyness and fear of making mistakes. In this case, even if teachers provide opportunities for speaking dialogue, students may choose to give up. Therefore, to solve this problem, educators need to reconsider their teaching methods. They need to pay more attention to students' actual speaking abilities rather than just mastering theoretical knowledge. Therefore, it is necessary to create an encouraging classroom environment in English speaking teaching where students are not afraid to make mistakes. Only in this way can students truly get speaking dialogue training to improve their speaking ability. The maturity and development of AI large language models(LLM) have facilitated the realization of such teaching practices[1,2].

Since the second half of 2022, with ChatGPT as the representative of LLM becoming a social focus, the introduction of various generative artificial intelligence technologies has provided

strong technical support for foreign language speaking auxiliary training and teaching[3,4]. The integration of AI models to achieve Automatic speech recognition(ASR) technology and text-to-speech (TTS) technology, combined with the use of AI knowledge bases, vector databases, and other technologies, will be an important direction for AI-assisted educational innovation[5,6]. The AI English speaking auxiliary teaching system built by our research team can provide personalized and interactive learning experiences for English speaking teaching activities by simulating real speaking environments.

AI Large Language Models can understand and generate human-like responses, allowing learners to practice conversational skills in a simulated and realistic environment. The intelligent system can instantly assess learners' pronunciation, grammar, vocabulary usage, and fluency, all of which are key components of effective communication. In addition, the AI assistant's adaptive learning ability ensures that the training difficulty and content can be personalized according to the current level of learners, thereby improving learning efficiency and targeting. In addition to core language skills, the AI system can also integrate cultural and situational learning to enrich the overall learning experience. By understanding the subtle differences in language use in different social and cultural backgrounds, learners can comprehensively understand the language and improve their ability to communicate effectively in a variety of situations. Using AI technology to create an immersive learning ecosystem can meet the personalized needs of learners. The potential of this technological reform for foreign language education is huge.

2. System Design and Construction

In this research and application project, we are committed to developing an innovative "Artificial Intelligence-Assisted English Speaking Teaching System" aimed at improving the efficiency and effectiveness of English speaking teaching through advanced AI and LLM technologies. The following is a detailed introduction to the system design and construction of our system:

2.1. System Design

In this study, the AI English speaking auxiliary teaching system we built mainly includes the following functional modules:

- (1). Intelligent voice recognition and feedback: The system uses Automatic Speech Recognition technology to accurately identify students' oral pronunciation and provide instant feedback to help students correct pronunciation errors[6]. In addition, the system will also use speech understanding technology to assess students' grammar and vocabulary usage, providing personalized learning suggestions.
- (2). Simulated dialogue environment: The system will use artificial intelligence technology to create a realistic dialogue environment, allowing students to practice simulated dialogues. Focusing on one-on-one oral dialogue training, the content of the dialogue is dynamically generated in real-time by the LLM platform configured with Prompt or pre-trained fine-tuning, ensuring that the dialogue content is close to that of a real person.
- (3). Adaptive learning path: The LLM platform configured with Prompt or pre-trained fine-tuning can automatically adjust the difficulty and content of oral dialogue according to students' learning progress and ability. The system can use machine learning algorithms to record students' learning situations and progress data, automatically feedback teaching progress and content, allowing each student to learn at their own pace, thereby improving learning efficiency.
- (4). Scenario-based interactive course learning: To improve students' learning interest, the AI auxiliary teaching system also introduces course scene interaction, that is, according to the teaching needs, AI adopts role-playing methods, allowing students to practice oral dialogue

with AI in simulated scenarios such as shopping, ticket booking, hotel registration, restaurant ordering, etc., increasing the authenticity and sense of presence of English training.

(5). Learning data tracking and analysis: The system will continue to track and record students' learning data, such as learning time, oral dialogue speed, sentence length, grammar assessment, etc. Through data analysis, teachers and students can clearly understand the learning progress and effects, so as to timely improve teaching methods and strategies.

2.2. System Construction

In the practice and research, we are committed to exploring and implementing an efficient and intelligent AI large model dialogue function module, aiming to provide learners with a smooth and natural English speaking practice environment. To achieve this goal, we have adopted a variety of advanced technologies and models to ensure the flexibility, accuracy, and user-friendliness of the system.

(1). Firstly, we have chosen the OpenAI ChatGPT4 API interface provided by Microsoft China as our main dialogue engine. ChatGPT4 is the latest iteration of Microsoft's large language model, which has strong natural language understanding and generation capabilities, and can engage in smooth conversations with users, providing accurate answers and suggestions. By accessing this API, the system can understand and respond to various oral practice needs of learners, thereby providing a high-quality dialogue experience. To further enhance the adaptability of the system, the project system also integrates the open-source large model Llama3 and its pre-trained model. To enhance the system's knowledge base support, we have also adopted the Retrieval-Augmented Generation (RAG) technology and integrated it into the dialogue generation process[7]. In this system, ChatGPT4 and Llama3 models can be switched on demand. The system can not only handle English speaking practice but also understand and respond to Chinese-related inquiries and requests. This technical integration allows the system to support as many issues as possible that learners may encounter in oral practice.

(2). The front-end dialogue user interface for learners uses WeChat mini-programs combined with Html5 in two ways to support the usage habits and preferences of different users. The WeChat end uses the WeChat simultaneous interpretation plug-in for real-time streaming voice recognition, implemented through the voice recognition manager RecoManager. The Html5 front end uses the NetClass H5 page voice recognition manager, allowing users to use this system on browsers that support Html5 without installing any additional plug-ins or applications. The text-to-speech conversion engine uses iFLYTEK's TTS3.0 voice engine, which can achieve voice output that simulates a variety of human voices.

(3). This system has also built a system application server and database to achieve digital management of teaching content, teaching records, and statistical analysis of teaching data, etc. The system application service makes the storage, distribution, and updating of teaching resources more efficient and convenient, and digital management ensures the traceability of teaching content. Teachers can adjust teaching plans and simulated scene content based on student feedback and learning progress. The system supports teachers in tracking the learning process and also supports students in checking their learning records at any time.

(4). In the evaluation of training effects, especially the effect evaluation of learners' oral English level, we adopt a combination of student subjective questionnaires and system objective record data. The relevant learning data recorded in the system's background, we mainly use sentence length[8] and speech rate[9] as a reference basis for the reference assessment of the system's English effect. In addition, in the system design, the dataset of sentence length and speech rate can be provided to the AI model for secondary analysis, and the system can use this as a basis for personalized teaching, providing dialogue practice of corresponding difficulty to students at different stages, allowing learners to be in the best learning area.

(5). The system's AI-assisted teaching-related data statistical analysis module has achieved data management of learners' learning records, learning habits, performance assessment, learning behavior data, etc., as well as teachers can set up scenario-based course management, class grouping management, and has built a data analysis and graphical display functions.

The following Figure 1 is a structural diagram of our AI-assisted English speaking teaching system.

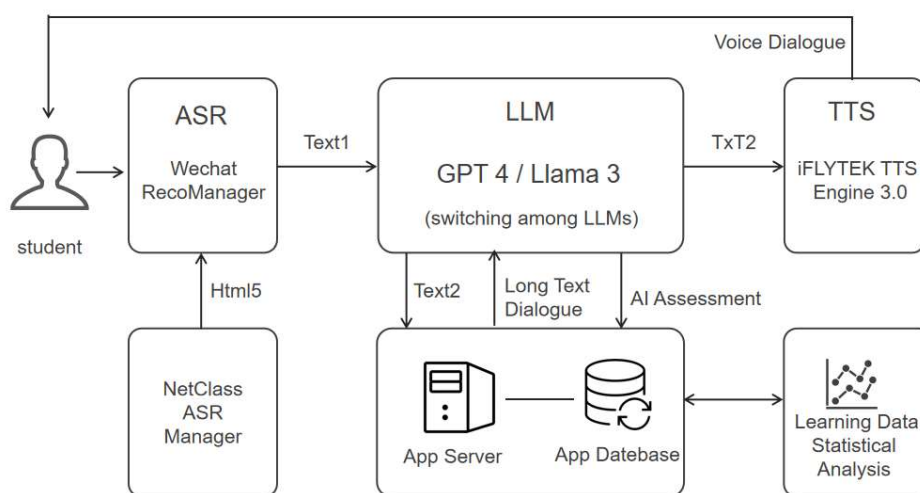


Figure 1. System Architecture Diagram

3. Results of Practice and Research

The AI-assisted oral English teaching system has been gradually piloted in various types of colleges and universities at Shanghai since August 2023. Most students used the WeChat interface to engage in simulated dialogue training with the AI oral teacher, conducting simulated dialogue training based on the scenarios set in college English courses. Both teachers and students reported achieving relatively good training and learning effects.

This paper uses case analysis methods to evaluate the effectiveness of AI technology in enhancing English speaking skills by collecting and analyzing platform usage data, student feedback, and teaching outcomes. At the same time, combining questionnaire surveys and interviews, we understand the acceptance and satisfaction of students and teachers with AI-assisted teaching. The combination of these two methods analyzes the practical effects of the AI English speaking assistance teaching system, and the results are organized as follows.

3.1. Learning Effectiveness Survey

A questionnaire survey was conducted on college students using the AI-assisted teaching system. A total number of 113 questionnaires were randomly distributed, and 113 valid questionnaires were collected. The questionnaire evaluated the students' self-assessment of the AI oral teaching system (AI oral teacher). The results are shown in Table 1 to 3.

Table 1. Learning Effectiveness Survey of AI System Improving English Speaking

Questionnaire Answers	Counts	Percentage
Definitely helpful	47	42%
Somewhat helpful	66	58%
Not helpful	0	0

From the questionnaire survey results, college students have a relatively high level of satisfaction with the AI-assisted oral teaching system. In actual use, students have a relatively equal preference for free dialogue training and scenario simulation dialogue training with the AI teacher. Generally speaking, at the beginning of system use, students prefer free dialogue, and later they use the course scenario dialogue mode more, which may be related to the requirements of foreign language teachers in the classroom.

Table 2. Training Mode Preference Survey

Questionnaire Answers	Counts	Percentage
Prefer free dialogue	8	7%
Prefer scenario simulation dialogue	43	38%
No bias	62	55%

The questionnaire survey results also show that the three aspects that students think are most effectively helped after using the AI-assisted training system are "helping themselves to speak boldly," "making their oral language more fluent through practice," and "providing their own listening skills." We need to pay attention to the answer to the first question in Table 3 of the questionnaire survey results. 82% of students answered "Helped me to speak up more boldly", indicating that AI platforms can help students overcome psychological barriers and achieve bolder oral expression and practice when learning English as a foreign language, without worrying about others' evaluations, which is a common problem in one-on-one foreign language training. This indicates that psychological factors in oral training are worth long-term attention and emphasis by teaching managers. In addition, the survey questionnaire in Table 3 also indicates that AI assisted dialogue training systems have significant benefits in improving English listening abilities.

Table 3. How the AI system improving Learner's English Speaking (multiple choice)

Questionnaire Answers	Counts	Percentage
Helped me to speak up more boldly	93	82%
Made my spoken language more fluent through practice	76	67%
Learned better ways to express myself in English	55	47%
Gained knowledge in a broader professional fields	39	35%
Improved my pronunciation by listening carefully	44	39%
Enhanced my listening skills	60	53%
Other	1	1%

In teaching practice, an important function of the AI system is to replace foreign language teachers to conduct one-on-one oral dialogue training with students, which is a difficult task for foreign language teachers in traditional teaching methods. In actual English oral teaching, one-on-one dialogue training and assessment between teachers and students are the most effective teaching methods but also the most time-consuming for teachers. Generally, it is difficult to achieve oral interaction between teachers and each student in the classroom, let alone one-on-one oral ability assessment by teachers for each student. The AI system can easily achieve the above functions, to a considerable extent, replacing foreign language teachers to complete the above work, and simultaneously achieve the recording and statistical analysis of teaching data. At the same time, students also have an opportunity for AI-supervised oral learning and can quantitatively evaluate their personal learning effects. In addition, since the AI teaching assistance system is deployed on personal mobile terminals, it objectively has a privacy

protection effect on students with lower oral levels, encouraging students to engage in more relaxed dialogue training with AI roles, thereby achieving good results.

3.2. Learning Data Sampling Analysis

In the evaluation of training effects, especially the effect evaluation of learners' oral English level, we adopt a combination of students' subjective questionnaires and objective record data of the system. The learning data recorded in the system's background, we mainly use sentence length and speech rate as a basis and serve as a reference for the evaluation of students' English oral ability progress after oral dialogue practice[8,9].

In this study, we randomly sampled the sentence length and speech rate data of student samples recorded in the AI-assisted oral teaching system's background teaching database to achieve a quantitative evaluation of the actual effects of AI-assisted oral practice. In practice, we define the beginning stage as the first 1 to 4 days of using the AI-assisted oral teaching system and the end stage as the last 4 days of completing the AI system practice.

First, we analyzed the change in oral sentence length. The ability to engage in long sentence dialogue is an important indicator of the improvement of oral expression ability. The data shows that after 10 to 30 days of training, most students' oral sentence length has significantly improved, with the average number of words per sentence increasing from 9.8 to 18.7 as show in Figure 2.

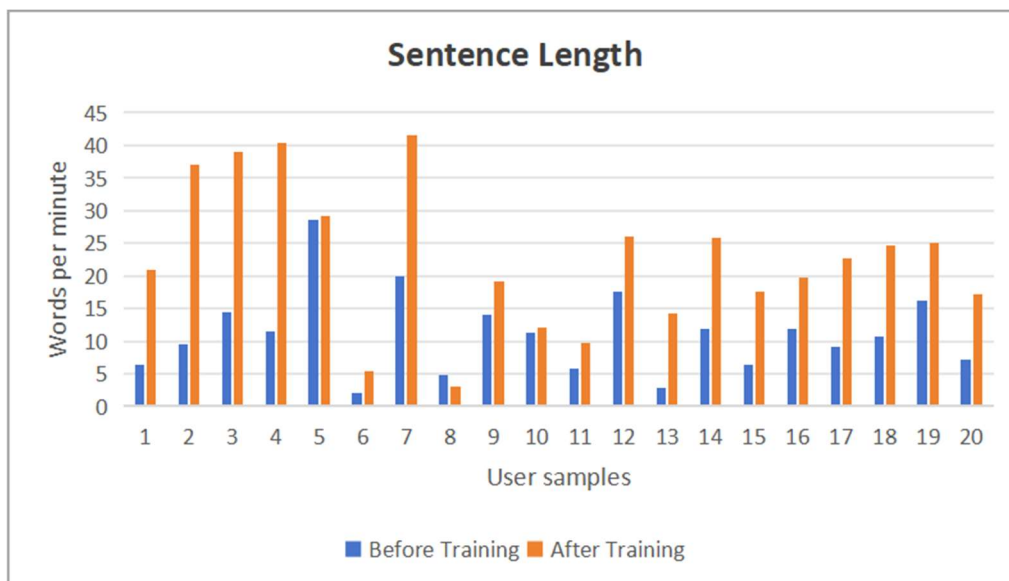


Figure 2. Comparison of the Sentence Length of Students Before and After AI Training

Secondly, we conducted a sampling statistical analysis of the changes in speech rate at the beginning and end stages. Speech rate is also an important reference indicator for examining English oral output ability. Generally speaking, fluent expression corresponds to a relatively stable speech rate. In this study, we define speech rate as the number of words per minute W/M. Data statistics on 143 students who participated in AI dialogue training for ten days to a month show that the average speech rate increased from 82W/M to 87W/M, and the data distribution is relatively average. As shown in Figure 3, we have randomly displayed the speech rate changes of 20 samples from a total of 143 students, with a comparison showing the contrast in speech rate before and after the training.

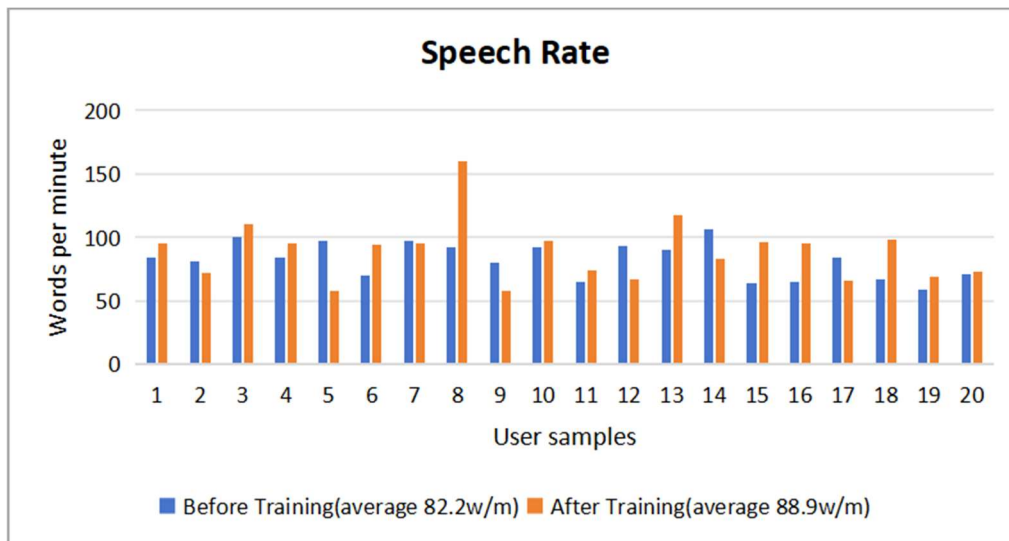


Figure 3. Comparison of the Speech Rate of Students Before and After AI Training

The research data shows that after practice with the AI oral assistance teaching system, most students' speech rates have increased. Corresponding to the students' English oral fluency, there is a significant improvement, further verifying that AI dialogue training has a positive effect on English oral output ability.

In teaching practice, we found that students are generally very interested in and willing to use the AI-assisted teaching system. At the same time, the encouragement and requirements of teachers for students are also very important factors, ensuring the effectiveness of students' dialogue training. For the improvement of students' foreign language oral ability, speaking dialogue itself has achieved teaching effects to a certain extent. Therefore, interesting and practical dialogue scene settings, combined with positive encouragement and teaching requirements, can give full play to the role of the AI-assisted teaching system and achieve innovative teaching results.

4. Conclusion

The application of AI technology, especially the use of large language models, has provided new solutions for college English speaking teaching. From this study and practice, it can be seen that AI technology has a relatively obvious effect in improving students' English speaking abilities. AI-assisted English teaching platforms can effectively help students learning English as a foreign language to quickly improve the fluency of their spoken English, and to practice English speaking more actively and boldly, thereby achieving significant learning progress. In future educational practices, we should continue to explore the deep integration of AI technology and teaching to promote the comprehensive improvement of students' English speaking abilities.

In addition, the introduction of AI technology also helps to cultivate students' autonomous learning ability and achieve the effect of classroom teaching. Students can practice speaking at any time and any place through the system, no longer limited by traditional classroom time and space. This flexibility allows students to better arrange their own study time and learn at their own pace.

In summary, the application and practice of AI technology in college English speaking training have achieved significant teaching effects. Besides, the application of AI systems has also improved the efficiency of teaching management and evaluation, and provided learners with a more personalized and interactive learning experience. By continuously optimizing and

improving AI technology, we can look forward to achieving more efficient, fair, and personalized language learning in the future.

References

- [1] G. L. Liu, R. Darvin, and C. Ma, (2024) 'Exploring AI-mediated informal digital learning of English (AI-IDLE): a mixed-method investigation of Chinese EFL learners' AI adoption and experiences', *Computer Assisted Language Learning*, p. 1–29.
- [2] X. Zhang, J. Sun, & Y. Deng. Design and Application of Intelligent Classroom for English Language and Literature Based on Artificial Intelligence Technology. *Applied Artificial Intelligence*, Vol. 37(2023) p. 1553-1575.
- [3] A. Baevski, S. Edunov, Y. H. Liu, L. Zettlemoyer, and M. Auli. 2019. Cloze-driven Pretraining of Self-attention Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, p. 5360–5369.
- [4] T. B. Brown, et al. Language Models are Few-Shot Learners. *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, Article No.: 159, P. 1877 - 1901.
- [5] R. Shadiey, W. Y. Hwang, N. S. Chen and Y. M. Huang. Review of speech-to-text recognition technology for enhancing learning. *Journal of Educational Technology & Society*, Vol. 17(4), p. 65–84.
- [6] D. Nickolai, E. Schaefer, and P. Figueroa, Aggregating the evidence of automatic speech recognition research claims in CALL, *System*, Vol. 121, p.103250-103260.
- [7] S. S. Manathunga, Y. A. Illangasekara. Retrieval Augmented Generation and Representative Vector Summarization for large unstructured textual data in Medical Education. Information on <https://doi.org/10.48550/arXiv.2308.00479>.
- [8] M. Darling-White and S. Banks. Speech Rate Varies With Sentence Length in Typically Developing Children. *Journal of speech, language, and hearing research*, Vol. 64(2021), P. 2385-2391.
- [9] R. L. Street and R. M. Brady (1982). Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communication context. *Communication Monographs*, Vol. 49(1982), p. 290–308.