International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-7

https://doi.org/10.70917/ijcisim-2026-0094
Article

Research on Interactive Learning Platform Based on Speech Recognition and Generation Technology in Higher Education English Classroom Teaching

Bing Shen *

School of International Education, Yellow River Conservancy Technical University, Kaifeng, Henan, 475004, China; hhsyshenbing@163.com

Abstract: The application of an interactive learning platform based on speech recognition and speech synthesis technologies in college English classrooms can facilitate effective speech-based interactive learning between teachers and students. This paper analyzes the use of interactive learning platforms in college English education to explore methods for enhancing learning outcomes. The teaching platform developed in this paper can receive and process speech data using speech recognition technology to assess the accuracy of students' pronunciation during reading aloud and provide feedback to learners. Additionally, speech synthesis technology is utilized to create authentic and diverse teaching scenarios for speech dictation, thereby enhancing students' listening comprehension skills. Furthermore, speech synthesis technology is used to help students “imitate” real people and “immerse” themselves in teaching scenarios and real-life teaching situations, thereby increasing students' in-class listening and speaking time and achieving constructive application of student learning evaluations. Research results show that the use of this teaching platform has improved human pronunciation accuracy from 76.5% to 91.2%, increased the number of speaking instances from 3.6 to 5.7, and significantly enhanced student participation and interest in learning. The platform can intelligently analyze learning trajectories and provide personalized feedback through interaction with individuals, serving teachers in developing differentiated instruction for students. By applying speech recognition and speech generation technology to analyze learners' pronunciation and provide feedback, this study assists English classrooms in achieving a new form of interactive education, which holds significant positive implications for the effective implementation of English education and teaching.

Keywords: speech recognition; generation technology; interactive learning; college English teaching; educational reform

1. Introduction

In recent years, with the advancement of information technology, artificial intelligence has gradually been applied to the field of education, particularly in the past few years, where speech recognition and speech generation technologies have been widely adopted in language instruction, driving significant changes and developments in university English classroom teaching models and methods [1-2]. Traditional English teaching models are constrained by factors such as classroom duration, teacher capabilities, and teaching methods, resulting in limited opportunities for students to engage in high-frequency, precise English speaking practice. As a result, students' speaking skills often fail to improve effectively [3-4]. Especially in large-class English settings, teachers struggle to provide individualized feedback on each student's pronunciation, severely limiting the effectiveness of speaking instruction. Additionally, low student engagement and a lack of effective communication in the classroom are unavoidable teaching phenomena, making it difficult to guarantee teaching outcomes [5]. English speech recognition technology can analyze and process real-time student speech data instantly,

while speech synthesis technology can create interactive language scenarios similar to real-life language environments. The combination of these two technologies provides immersive, interactive, and personalized technical support for foreign language learning [6-7]. Students can use intelligent platforms to practice and test their English speaking skills anytime and anywhere, while teachers can analyze the data generated by the system to achieve targeted teaching and precise measurement [8-10]. Therefore, utilizing speech technology and speech synthesis technology to construct an interactive foreign language learning platform for teachers and students is an important way to achieve the transformation of foreign language teaching from teacher-centered to student-centered and to implement the English classroom revolution in higher education, thereby achieving its educational and teaching objectives.

This paper addresses issues such as delayed feedback, limited interaction, lack of learning specificity, and monotony in traditional English classroom activities by designing an interactive classroom teaching and learning system based on speech recognition and synthesis technology. This system enables real-time speech input and collection, with the processed results outputting corresponding English learning content and learning outcomes for learners. It also evaluates, corrects, and provides feedback on students' language output, and pushes learning feedback reports to students, further enhancing their adaptive learning capabilities. Based on this design, the learning system was tested through experiments comparing English classroom learning feedback and teaching efficiency under both platform and non-platform modes to validate the platform's teaching applicability.

2. Research Review

In recent years, there has been an increase in research on interactive English teaching and learning, with many scholars dedicated to developing an excellent interactive English learning platform. Reference [11] developed an online interactive teaching platform tailored for university English proficiency courses, enhancing the richness of interactive content. The platform was optimized using cloud service technology. Reference [12] created an intelligent learning system for university English interactive learning. The system integrates multiple modules and is designed based on principles such as interactive learning, system interface, and learning functionality. Test results indicate that the system effectively improves students' learning efficiency. Literature [13] designed an interactive practice platform supported by feedback-based elite teaching optimization algorithms and open API protocols for university English mobile teaching environments. The platform has a response time as low as 0.85 seconds, with optimized CPU and memory usage, providing users with a better practice experience. Literature [14] created an interactive English teaching software with four major modules (management and teaching settings, multiple interactive English teaching, exam paper distribution, and database), which was effectively used in experiments involving speaking, listening, and reading. Literature [15] reports that active learning assisted by the interactive learning platform ClassPoint can improve students' memory of English vocabulary, but it is not ideal for mastering grammar. With the development of speech recognition and speech synthesis technology, attention has been drawn to the field of English education. Literature [16] provides an interactive application based on language teaching and learning theory, computer-assisted language learning, speech recognition technology, human-computer interaction, and scaffolding teaching, designed for English speaking practice. Literature [17] found that although AI-synthesized speech lacks fluency, it qualifies as high-quality speech coding. AI speech synthesis technology and wireless network technology-generated audio outperform textbook audio in terms of expression and can help students achieve higher listening scores.

3. Research Methods for Interactive Learning Platforms Using Speech Recognition and Synthesis Technologies

3.1. Study Design

The article designs a teaching platform system for teacher-student interaction based on speech recognition and speech synthesis. The technical system established in the design adopts a three-layer architecture comprising the data layer, business layer, and interaction layer. The data layer stores student learning activities, speech, and textbook data; the business layer implements speech recognition and synthesis; and the interaction layer implements the data architecture design for teacher-student interaction. Under the designed process, a system architecture comprising six functional modules was completed, as shown in Table 1. Speech recognition is achieved using deep learning technology to identify students' speech input content and perform real-time speech analysis and processing. Speech synthesis technology provides teacher-standard speech expression output examples for students' speech output.

Table 1. The design of core functional modules of the Interactive Learning platform.

Module name	Function description
Speech recognition	Real-time voice input recognition, pronunciation accuracy assessment, and voice transcription function.
Voice generation	Standard voice generation, situational dialogue simulation, and voice prompt feedback
Interactive practice	Oral practice tasks, online voice conversations, and classroom interactive Q&A
Learn to analyze	Learn data collection, progress tracking and analysis, and personalized report generation.
Resource management	Teaching resources are uploaded, classified and managed, and shared and distributed
System management	User permission management, system configuration maintenance, data backup and recovery.

From a spatial perspective, online platforms break the physical constraints of traditional classrooms on the time and space of teachers and students, enabling real-time teaching and learning for everyone. From a relational perspective, they redefine the relationships between teachers and students, as well as among students themselves, shifting from the traditional teacher-centered model to one where teachers act as facilitators of learning. Online teaching platforms offer a variety of courses, diverse types of instructional content, and expanded platforms, collecting and managing resources automatically generated and uploaded by teachers. Using responsive development, the interface is designed to be multi-device compatible, enabling users to view visualizations of their learning progress data. Standard data interfaces are reserved for integration with other campus systems, and based on a modular development architecture, interfaces are reserved for future functional modules.

3.2. Experimental Process

The author designed and implemented a comparative study on the educational impact of an interactive learning platform based on speech recognition and synthesis technology. A total of 60 second-year English majors from a certain university were selected, and based on their specific circumstances, they were divided into two groups using a lottery method: the experimental group and the control group. Students in the experimental group used an interactive learning platform with speech recognition and synthesis functions for a period of eight weeks, while students in the control group learned through traditional classroom instruction. During the study, the researcher made efforts to select a sample that could broadly represent the research subjects, striving for consistency in terms of gender and English proficiency to minimize the error rate among the research subjects. To ensure the smooth conduct of the experiment, the experimental group received one week of training in the use of the interactive learning platform prior to the 8-week experimental period, enabling them to become proficient in operating the speech recognition and synthesis functional modules.

During the experiment, to effectively track the development and changes in language abilities between the experimental and control groups, a post-experiment qualitative analysis was conducted using a testing method. The results of oral and listening ability tests were used to assess the participants' progress in foreign language learning. A satisfaction survey questionnaire based on the Likert five-point scale was developed to assess participants' satisfaction levels with different teaching methods. Additionally, a classroom behavior observation survey form was used to conduct qualitative analysis of participants' engagement, activity levels, discussion participation, and learning enthusiasm during class. During the process of statistical data and information organization, a combination of qualitative and quantitative analysis was employed to statistically analyze differences in the learning process and learning outcomes of the participating students. The effectiveness of the interactive teaching platform compared to traditional methods in enhancing foreign language teaching outcomes and its role in the foreign language learning process of students was evaluated.

3.3. Data Analysis Methods

This paper employs a multi-level data analysis approach, combining quantitative and qualitative methods to comprehensively evaluate the educational effectiveness of an interactive learning platform. Data analysis is conducted across multiple dimensions, including the performance of speech recognition technology, learning outcomes, and classroom engagement levels. The data processing process is standardized, with raw data collected by the platform undergoing cleaning, coding, and classification,

and outliers and missing data being removed to ensure the accuracy of subsequent statistical analyses. The data analysis indicators are subject to multiple validations, with results from different algorithms compared to ensure reliability. The speech recognition accuracy rate of the platform is a key technical performance indicator for interactive learning platforms, calculated using the formula:

$$Accurary = \frac{Number\ of\ Corrent\ Recognitions}{Total\ Number\ of\ Recognitions} \quad (1)$$

In the formula, the number of correct recognitions and the total number of recognitions represent the number of times the speech recognition system correctly identifies speech input and the total number of all speech inputs, respectively. This can be used to measure the recognition accuracy of the recognition system. In this paper, the recognition accuracy rate is categorized into different groups based on factors such as whether the speech is clear, fluent, or accented. This allows for a comparison of recognition accuracy rates under different conditions, highlighting the strengths and weaknesses of speech recognition and providing guidance for targeted technical improvements in the future. The learning effect growth rate is calculated by subtracting the pre-test score from the post-test score, with the formula being:

$$Improvement\ Rate = \frac{Post\_test\ Score - Pre\_test\ score}{Pre\_test\ score} \quad (2)$$

This metric effectively reflects the extent of changes in students' academic performance before and after learning on the interactive platform. To make learning outcomes more transparent, we have broken down academic performance into multiple dimensions such as oral skills, listening skills, and pronunciation skills, and calculated the improvement rate for each dimension separately, thereby clarifying the platform's effectiveness in specific language skills.

Additionally, a learning efficiency metric is constructed, using the increase in learning progress per unit of time to measure the platform's teaching efficiency. Classroom participation analysis is an analytical method combining behavioral observation and data mining, constructing a participation analysis model based on behavioral characteristics such as the number of times students speak during class, the number of questions asked, and group collaboration and communication. The platform records students' learning behavior data, such as login frequency, exercise completion status, and interaction response time, which objectively reflect participation levels. Furthermore, correlation analysis methods are employed to analyze the relationships between variables, identifying key factors influencing learning outcomes, thereby providing deeper insights for future educational practices. All data analysis processes and results effectively support the practical application of interactive learning platforms.

4. Teaching Results and Performance Test Analysis

4.1. Teaching Experiment Results

Through statistical analysis of the teaching experiment data over an eight-week period, a comparison of the performance between the experimental group and the control group revealed the following results, as shown in Table 2. In terms of English learning, the interactive learning platform utilizing speech recognition and generation technology demonstrated positive outcomes. Students in the experimental group showed a more significant improvement in English listening skills compared to the control group. The average score on the final exam increased by 17.3 points from the initial exam, while the control group's average score was only 73.2 points, with an increase of just 8.9 points. The speech recognition accuracy rate of the experimental group also improved significantly, rising from 76.5% to 91.2%, further demonstrating that the use of an interactive learning platform can enhance students' pronunciation accuracy.

Table 2. The experimental group was compared to the control group.

Test items	Experimental group		Control group	
	Before	After	Before	After
Listening comprehension	68.3	85.6	67.9	73.2
Oral expression	65.7	83.4	66.1	72.8
Speech accuracy	76.5%	91.2%	75.8%	82.3%
Comprehensive score	70.2	86.7	69.8	76.1

There is significant variation in students' participation in classroom teaching. Figure 1 illustrates the correlation between students' in-class participation in learning and their academic performance, while Table 3 presents the in-class participation in learning and satisfaction with classroom learning across

different classes. As shown in Figure 1, students with higher levels of in-class participation tend to achieve better English grades. The interactive learning platform designed in this study can enhance students' participation in classroom teaching, resulting in significantly higher English grades compared to the control group. The experimental results indicate that in classes using the interactive learning platform, the average classroom participation time per class ranges from 2.5 to 5.7 times, with 92.3% participation in group discussions. This level of active participation is difficult to achieve in traditional teaching methods. In contrast, the control group students had an average classroom participation time of only 2.3 to 2.8 times per class, with 67.5% participation in discussions. A student satisfaction survey conducted using the Likert five-point scale showed that the experimental group rated the platform with an overall score of 4.6, while the control group rated it with an overall score of 3.4. Students in the experimental group reported increased interest in learning and significantly longer post-class self-study time.

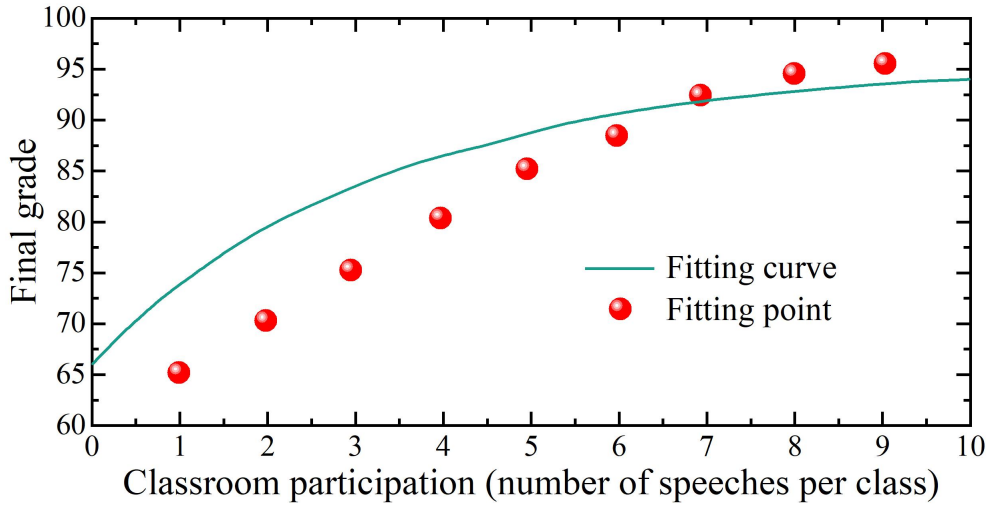

Figure 1. Students' classroom participation and grades.

Table 3. Class participation and satisfaction scores.

Test items	Experimental group		Control group	
	Before	After	Before	After
Speaking number	2.5	5.7	2.3	2.8
Group discussion engagement	64.1	92.3	64.7	67.5
Satisfaction score	3.2	4.6	3.1	3.4

In summary, it is evident from the above experimental research that the interactive language teaching platform offers significant advantages in enhancing the quality of English learning and classroom language communication skills among college students. The experimental group demonstrated substantial improvements in listening, speaking, and pronunciation, and the platform's real-time feedback and targeted language learning guidance received unanimous praise from students. Most students in the experimental group felt that the learning platform had a positive impact on improving their English learning, which also fully demonstrated the innovative value and significance of interactive teaching methods in the reform of English teaching in Chinese higher education institutions.

4.2. Platform Optimization Testing

This paper focuses on user research and designs a learning system based on voice interaction using a multi-layer architecture. The front end is developed using a responsive framework, and the system's functionality is implemented through the iFlytek voice technology platform. A front-end/back-end separation design approach is adopted, with data transmission between layers—including the data storage layer, business logic layer, and user interface layer—facilitated through standardized interfaces.

In terms of technical implementation, the speech recognition module utilizes the multi-head attention mechanism from deep learning, while speech synthesis employs a sequence-to-sequence neural network structure, ensuring high accuracy in recognition and naturalness in generated speech. The system optimizes concurrent processing capabilities through a distributed implementation scheme and data

caching, and employs token authentication to ensure user data security. The effectiveness of the design optimization methods for the system described in this paper is validated through analysis. The system's response time is primarily validated through the number of concurrent users, with the changes in response time before and after optimization shown in Figure 2. As shown in Figure 2, when the number of concurrent users reaches 1,000, the response times before and after optimization are 1,976.42 ms and 1,011.36 ms, respectively, with the optimized system's response time reduced by 48.83% compared to the pre-optimization time. Through extensive testing and optimization, the system can handle high concurrency scenarios, with response times consistently meeting expected targets.

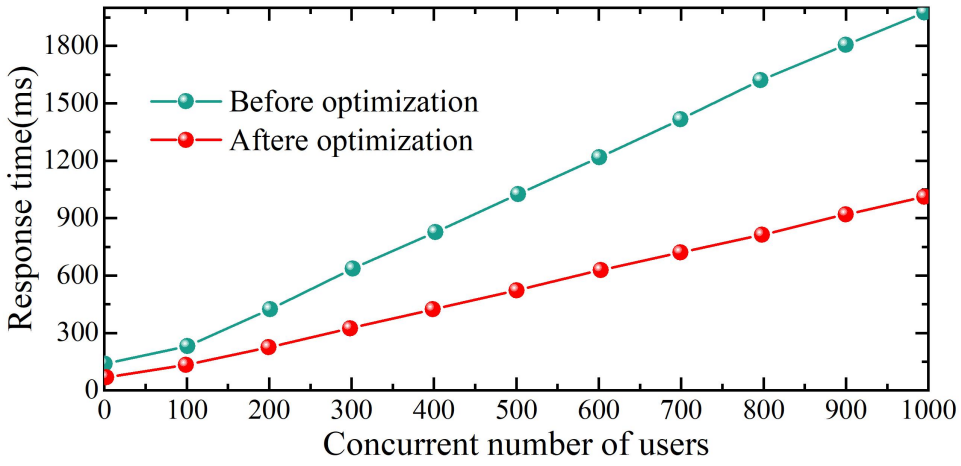

Figure 2. Platform optimization testing.

5. Conclusion and Outlook

5.1. Conclusion

Based on the demand for reforming English teaching models in higher education institutions, the author conducted research on interactive learning platforms enabled by speech recognition and speech synthesis technologies. The findings confirm that the technological capabilities of interactive learning platforms can effectively promote teaching model reforms. The results indicate that interactive learning platforms not only transform traditional classroom teaching models by providing students with personalized speech correction and learning feedback but also significantly improve students' pronunciation accuracy rates, which increased from 76.5% to 91.2% during the initial learning phase. In classroom teaching activities, students' participation levels have significantly increased, rising from an average of 2.5 activities per class in the previous stage to 5.7 activities per class in this stage. The average overall satisfaction score for most students was 4.6 out of 5, with the majority indicating a significant increase in learning interest. This demonstrates the practical significance of interactive learning platforms in promoting the implementation of university English teaching.

First, from the perspective of the transformation of the teacher's role, the development of the Learning Pass platform has facilitated the transition of teachers from traditional lecture-based teaching to instructional guidance, and from being the primary focus of teaching to becoming implementers and controllers of the teaching process. The Learning Pass platform's teaching methods have broken through the temporal and spatial limitations and rigidity of traditional teaching, making instruction more targeted and flexible, with real-time feedback on learning progress. Additionally, the system enables real-time recording of teachers' instructional behaviors during the teaching process and provides feedback on students' learning progress and status, thereby allowing adjustments to teaching strategies. Finally, in the context of English instruction in an information technology environment, it is evident that with the integrated development of information technology, the evaluation methods in English instruction are shifting from traditional approaches to more objective and rational process-based evaluation methods. This establishes a framework and reference direction for future evaluation methods, making teaching reforms more advanced and laying the foundation for future higher education English instruction.

5.2. Outlook

Based on the practical process of this study, it is evident that there are still areas worthy of further exploration regarding the application of speech recognition and speech synthesis technologies in English

learning at higher education institutions. While our research has demonstrated the significant advantages of these technologies in teaching, current speech recognition and speech synthesis technologies still exhibit certain inaccuracies in recognition accuracy due to variations in classroom environments and accents. These issues necessitate further research to enhance the performance of technical algorithms and improve the adaptability of these technologies to diverse speech patterns. Given the aforementioned application scenarios, the high costs associated with the development, application, and maintenance of online interactive learning platforms have become a barrier to the widespread adoption of this teaching model in more universities. This necessitates proactive exploration of open-source technology applications, integration of cloud services, and shared resource-based teaching to reduce platform construction and operational costs, thereby enabling more universities to benefit from this new teaching model.

From the perspective of optimizing teaching, I believe there is significant room for research in enhancing students' autonomy and technological acceptance, thereby exploring how to design more engaging learning tasks to encourage student participation. From a technological standpoint, when utilizing and developing new technologies such as VR and AR, as well as multimedia applications to create immersive learning experiences, how to stimulate student interest and facilitate student interaction and collaboration through social media and online learning platforms is a new area of research. Further technological innovations should be pursued in areas such as technical optimization, cost reduction, enhancing students' learning initiative, and optimizing new media applications, to better enable students to utilize speech recognition and generation technology for English semantic expression.

References

1. Wei, L. (2019). Study on the application of cloud computing and speech recognition technology in English teaching. *Cluster Computing*, 22(Suppl 4), 9241-9249.
2. Zhang, Z. (2022). Application of intelligent speech synthesis technology assisted by mobile intelligent terminal in foreign language teaching. *Mathematical Problems in Engineering,* 2022(1), 9751094.
3. Tang, S., Long, M., Tong, F., Wang, Z., Zhang, H., & Sutton-Jones, K. L. (2020). A comparative study of problem-based learning and traditional approaches in college English classrooms: Analyzing pedagogical behaviors via classroom observation. *Behavioral Sciences*, 10(6), 105.
4. Tajenova, S. (2025). Evaluating Traditional And Contemporary Methods Of English Instruction In Academic Settings. *International Journal of Artificial Intelligence*, 1(2), 1888-1891.
5. Arifani, Y., Suryanti, S., Wicaksono, B. H., Inayati, N., & Setiawan, S. (2020). EFL Teacher Blended Professional Training: A Review of Learners' Online and Traditional Learning Interactions Quality. *3L: Southeast Asian Journal of English Language Studies*, 26(3).
6. Wang, J. (2020). Speech recognition of oral English teaching based on deep belief network. *International Journal of Emerging Technologies in Learning (Online)*, 15(10), 100.
7. Kenoui, M., & Ait Mehdi, M. (2020, May). Teach-me dna: an interactive course using voice output in an augmented reality system. In *2020 1st International conference on communications, control systems and signal processing (CCSSP)* (pp. 260-265). *IEEE.*
8. Zhang, W., Wang, B., & Zhou, L. (2022). Evaluation and Improvement of English-Speaking Instruction Based on PLS-SEM and Intelligent Speech System. *Mobile Information Systems*, 2022(1), 1081327.
9. Ren, Z. (2025). Design and Application of a College English Online Learning Platform. *International Journal of High Speed Electronics and Systems*, 2540442.
10. *Li, X. (2024). A Personalized Teaching System for College English Based on Big Data and Artificial Intelligence. Scalable Computing: Practice and Experience, 25(6), 5477-5485.*
11. Fan, L. (2022, September). Online interactive platform for college english intensive reading teaching based on cloud service. In *International Conference on Advanced Hybrid Information Processing* (pp. 467-480). *Cham: Springer Nature Switzerland.*
12. Wang, Y. (2018, August). Development and Application of An Intelligent Learning System for Interactive College English. In *2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)* (pp. 39-42). *IEEE.*
13. Wang, L., Jia, X., Cui, H., & Zhang, B. (2022). An interactive practice platform of English mobile teaching in colleges and universities based on open API. *International Journal of Continuing Engineering Education and Life Long Learning*, 32(4), 418-431.
14. Hou, X., & Zhang, Z. (2022). Multi-interaction English teaching platform based on internet of things. *International Journal of Continuing Engineering Education and Life Long Learning*, 32(4), 488-505.
15. Tiansoodeenon, M., & Prasongngern, P. (2025). Enhancing Active Learning through the Interactive Learning Platform to Improve Thai EFL Students' English Vocabulary, Grammatical Retention, and Motivation in English Learning. *Higher Education Studies*, 15(1), 232-244.
16. Oh, E. Y., & Song, D. (2021). Developmental research on an interactive application for language speaking practice using speech recognition technology. *Educational Technology Research and Development*, 69(2), 861-884.
17. Yu, C., Wu, L., Li, J., & Li, S. (2023). English Listening Teaching Mode under Artificial Intelligence Speech Synthesis Technology. *ACM Transactions on Asian and Low-Resource Language Information Processing*.