Journal of the Brazilian Computer Society, 2025, 31:1, doi: 10.5753/jbcs.2025.5089

This work is licensed under a Creative Commons Attribution 4.0 International License.

Training of English Listening and Speaking Using Virtual Reality Technology

Qinghua Yang [**Hubei University of Science and Technology** | ***yqh_yang@outlook.com***]

School of Foreign Languages, Hubei University of Science and Technology, Xianning 437100, China

Received: 18 October 2024 • **Accepted:** 14 May 2025 • **Published:** 30 August 2025

Abstract This paper briefly introduces virtual reality (VR) technology and the basic application process for English listening and speaking training. A case study was conducted with freshmen from the School of Foreign Languages at Hubei University of Science and Technology. The long short-term memory (LSTM) algorithm used for scoring spoken English levels was analyzed and compared with the traditional back-propagation neural network (BPNN) and recurrent neural network (RNN) algorithms. The students were divided into a control group and an experimental group for five-week training. The English listening and speaking levels of these students were tested before and after the training. The results showed that the LSTM algorithm had the highest accuracy and efficiency in scoring spoken English. After the training, the English speaking and listening scores of the students in the experimental group, which used VR technology, were significantly improved.

Keywords: Virtual reality, English, Speaking, Teaching Mode

1 Introduction

Although traditional methods of English listening and speaking training have their value, they often struggle to achieve the desired teaching effect due to factors such as teaching environment, faculty resources, and student engagement [Xie *et al.*, 2022]. Virtual reality (VR) technology allows users to immerse themselves in artificially created environments that simulate real or virtual scenes [Pack *et al.*, 2020]. When applied to English listening and speaking training, VR technology can create diverse English communication scenarios for students to assist in English listening and speaking training [Ma, 2021]. [Luo *et al.*, 2023] developed a spherical video virtual reality (SVVR) technology-based learning system and conducted a quasi-experiment in an English major's classroom at a university. The results indicated that the SVVR-supported English writing teaching method could enhance students' writing proficiency, enrich the content of their writing, and boost motivation for learning. [Wen and Fu, 2021] explored the use of VR technology to simulate an English teaching course for occupational health in higher vocational education and employed constructivism to create an authentic language environment. [Jiang *et al.*, 2021] constructed a novel English teaching system based on VR and evaluated its effectiveness. This paper briefly introduces VR technology and its application process in English listening and speaking training, followed by an analysis of a case involving freshmen from the School of Foreign Languages at Hubei University of Science and Technology. The limitation of this article lies in that it only uses an intelligent algorithm and the VR platform to conduct short-term training for freshmen. On the one hand, the number of students participating in the training was small. On the other hand, the duration was short, making it difficult to confirm the universality and durability of this learning method. Therefore, the future research direction is to increase the number of participants and extend the test duration. The contribution of this article lies in the utilization of VR technology and intelligent algorithms to assist students in the training of English listening and speaking, providing an effective reference for improving students' English proficiency.

2 English Learning Based on Virtual Reality Technology

The key of VR technology includes 3D modeling technology, interaction technology, and display technology [Duffy *et al.*, 2024]. 3D modeling technology involves converting natural objects, scenes, and characters into digital models within a computer. Interaction technology utilizes sensing devices to capture user movements, gestures, sounds, and other information [Wu and Qiu, 2022], which are then translated into digital signals, and the virtual environment gives feedback to the digital information. Display technology visualizes the virtual environment and its corresponding feedback [Zhou, 2020].

Traditional English listening and speaking training typically takes place in a classroom setting and is usually performed by the teacher. However, this conventional learning method is influenced by factors such as the teaching environment, teacher capabilities, and other constraints. For instance, the fixed location of the classroom and limited course time can impact the effectiveness of the training. Additionally, due to a limited number of teachers [Uruthiralingam and Rea, 2020], English instruction often involves one-to-many teaching, which can hinder the teacher's ability to focus on individual students and develop personalized teaching programs. This lack of individual attention can also impede student engagement

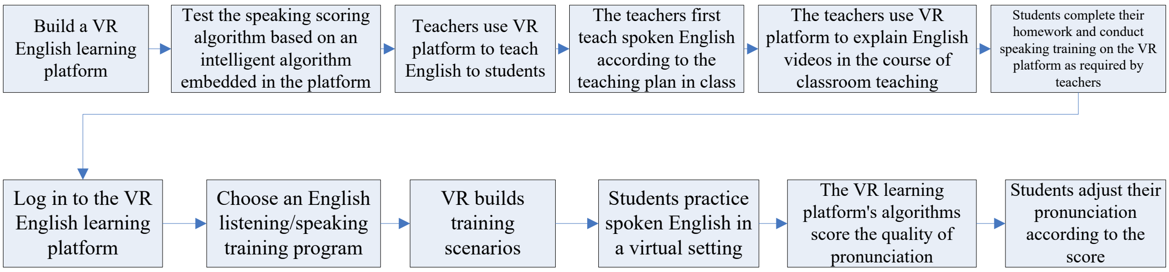
Build a VR English learning platform
Test the speaking scoring algorithm based on an intelligent algorithm embedded in the platform
Teachers use VR platform to teach English to students
The teachers first teach spoken English according to the teaching plan in class
The teachers use VR platform to explain English videos in the course of classroom teaching
Students complete their homework and conduct speaking training on the VR platform as required by teachers
Log in to the VR English learning platform
Choose an English listening/speaking training program
VR builds training scenarios
Students practice spoken English in a virtual setting
The VR learning platform's algorithms score the quality of pronunciation
Students adjust their pronunciation according to the score

Figure 1. Basic flow of English speaking/listening teaching mode based on VR technology.

and participation in the learning process. As English is fundamentally a language for communication, proficiency in listening and speaking is crucial. VR technology has the potential to offer students a realistic English communication environment and personalized learning experiences [Shang, 2022].

The basic process of the English speaking/listening teaching mode combined with VR technology is shown in Figure 1.

① Universities or teachers build a VR English learning platform by using the existing VR software. During the building process of this platform, an English speaking scoring algorithm based on an intelligent algorithms is embedded [Ding and Qi, 2022]. This algorithm can score students' spoken English, and its principle steps are pre-processing of audio, extracting audio features, forward calculation of the deep learning algorithm, and giving the score of pronunciation quality. The main purpose of pre-processing is to reduce the noise in the audio. The audio feature extraction adopts the mel-frequency cepstral coefficient (MFCC) feature that is closer to human ear discrimination [Han, 2022]. The deep learning algorithm used for the scoring of pronunciation quality adopts the long short-term memory (LSTM) algorithm, which is improved from the recurrent neural network (RNN), and a gating mechanism is introduced on the basis of a RNN, thereby avoiding the problem of gradient explosion or vanishing when RNN faces long sequence data [Li and Xie, 2021].

② After the VR English learning platform is built, the embedded spoken English scoring algorithm is tested to verify that whether the algorithm can accurately evaluate students' spoken English proficiency.

③ Teachers teach English to students using the VR platform. Firstly, they conduct regular spoken English teaching in the classroom according to the teaching plan. Meanwhile, during this process, teachers also use the VR platform to explain English videos to students and utilize the novelty of VR technology to attract students' attention.

④ After class, students complete the homework assigned by the teacher and conduct spoken English training on the VR platform as required by the teacher.

⑤ Before conducting speaking English training on the VR platform, students first log in to the VR English learning platform with their accounts.

⑥ Then, students choose the speaking English training items according to the requirements of the teachers or their interests and hobbies. The VR platform constructs the training scenarios based on the items selected by the students.

⑦ Students conduct training in the virtual scenarios. During the process, the VR platform provides dialogues that are in line with the virtual situations.

⑧ The English conversations conducted by students in the virtual scenarios are collected by the platform, and then the collected speech is scored using the embedded speaking English scoring algorithm.

⑨ Students eventually adjust their pronunciation based on the speaking English score given by the VR platform.

3 Case Study

3.1 Subjects

In this study, freshmen from the School of Foreign Languages at Hubei University of Science and Technology were divided into a control group and an experimental group for analysis. The experimental group used VR technology to support English listening and speaking training, while the control group relied on traditional multimedia teaching for the same training. The students have signed informed consent forms.

3.2 Experimental design

3.2.1 Performance test of the spoken English scoring algorithm

Before formally evaluating the impact of the VR technology platform on English-speaking training, the effectiveness of the spoken English quality scoring algorithm within the platform was initially tested. The algorithm's relevant parameters are as follows: the MFCC feature dimension was set to 12; the LSTM input layer comprised 12 nodes; the hidden layer had 64 nodes, the sigmoid activation function was used; the output layer had one node; the training was performed 400 times; the learning rate was 0.02; the stochastic gradient descent method was used.

Independently collected data on students' spoken English was used. The students' spoken English level was evaluated on a 10-point scale by five teachers with over five years of teaching experience. Sixty percent of the gathered samples were allocated to the training set, while the remaining 40% were designated as the test set. In addition, the traditional BPNN and RNN algorithms were also tested for comparison.

3.2.2 Influence of VR technology on the effect of English speaking and listening training

The VR platform for students' English listening and speaking tests is a simulation platform developed independently using VR software to create various conversation scenarios. Varjo VR-3 glasses were used. In this study, the VR platform was utilized to construct the business environment of a convenience store. All item labels were presented in English, dialogues with non-player characters (NPCs) in the scene were conducted solely in English, and other prompts were also in English, except for the Chinese prompts in the login interface. Moreover, the VR platform was integrated with the spoken English quality scoring algorithm proposed in this paper. During student training in the virtual scene, the VR peripheral captured their pronunciation, and the algorithm subsequently evaluated its quality. The scoring results were then displayed to the students within the virtual scene.

Prior to the English speaking and listening training, students in both the experimental group and the control group underwent assessments of their English speaking and listening proficiency. Then, the English speaking and listening training for both groups spanned five weeks, with three training sessions per week. Upon completing the training period, students in both groups were reevaluated for their English speaking and listening proficiency levels.

The control group adopted the traditional teaching mode, and the experimental group adopted the VR technology-based teaching mode. The traditional teaching mode included the following two content. ① In the ordinary classroom teaching, teachers conducted spoken English teaching for students according to the teaching progress. To be specific, the teacher read the textbook once, and the students followed. Then, relevant audio was played, and the students continued to follow. In addition, teachers also organized inclass tests, i.e., playing an English audio, and students answered the questions based on the audio. After class, teachers assigned one or more English audios and supporting questions. ② After students completed the homework, the teacher explained the difficult points in the questions in class.

The content of the VR teaching mode included four content. ① In classroom teaching, teachers also conducted spoken English teaching to students according to the teaching progress. Besides the traditional follow-up reading, teachers also used the VR platform to analyze the key points of the spoken English textbooks. ② During the classroom teaching process, teachers took out a period of time to let students watch an English video (including movies, TV series, etc.) on the VR platform and explained the key points of the language. ③ After class, in addition to the regular spoken English homework, students also discussed the difficulties of homework with classmates and teachers in the communication group. ④ In spare time, teachers asked students to conduct situational simulation dialogues in the VR platform to promote students' language sense.

After the training was completed and the listening proficiency was tested, a questionnaire survey was conducted for the students. The questionnaire was anonymous, and the questions included: ① What is your main motivation for learning on the VR platform: Required by the teacher; To consolidate knowledge; Pure interest? ② Which learning form do you prefer in VR teaching: Traditional classroom; VR platform? ③ Your communication status with others in the VR teaching mode: More active online; More active offline; Active both online and offline? ④ Do you think the VR teaching mode can meet your learning needs? ⑤ Which part are you more satisfied with in the VR teaching mode: Traditional classroom; VR platform?

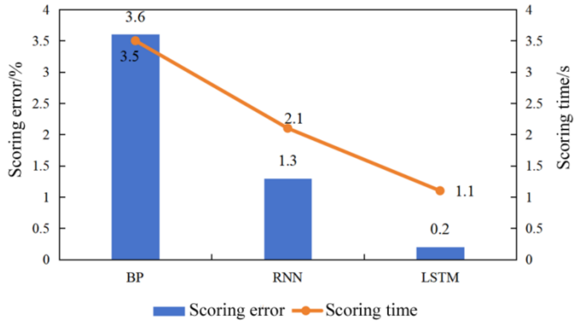

Figure 2. Performance of three spoken English quality scoring algorithms.

3.3 Mathematical statistics

To assess the impact of VR technology on the effectiveness of English speaking and listening training, statistical analysis was conducted using SPSS software [Guo and Gao, 2022] to analyze the English speaking and listening scores before and after the training.

3.4 Test results

Before using the VR technology platform for training students' English speaking and listening skills, the performance of three spoken English quality scoring algorithms was initially evaluated. The results of the performance analysis of the three algorithms are illustrated in Figure 2. The scoring error of the BPNN algorithm was 3.6%, with a scoring time of 3.5 s; the RNN algorithm had a scoring error of 1.3% and a scoring time of 2.1 s, while the LSTM algorithm showed a scoring error of 0.2% and a scoring time of 1.1 s. It can be seen that the LSTM-based scoring algorithm demonstrated the lowest scoring error and required the least time for scoring compared to the other algorithms.

The control and experimental groups underwent five weeks of spoken English training. The control group adopted the traditional multimedia teaching method, while the experimental group employed the VR simulation environment-based teaching method. The experimental group was assessed for the change in spoken English proficiency throughout the training process using the quality scoring algorithm. In contrast, the control group did not incorporate the scoring algorithm in their training. However, to compare the progress of the speaking level of both groups, this study also evaluated the change in the speaking level of the control group using the speaking scoring algorithm. The results are depicted in Figure 3. It can be seen that as the

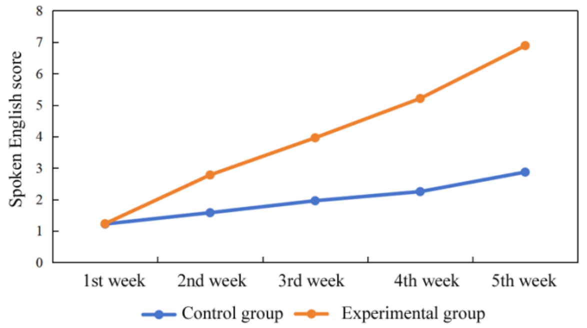

Figure 3. Changes in spoken English levels of the control and experimental groups over five weeks.

training duration increased, the spoken English proficiency of both groups of students improved, and a more significant enhancement was observed in the experimental group.

Table 1 shows no significant difference in English listening and speaking scores between the two groups before the training. Following the training, the control group's score exhibited an improvement, although not significantly. In contrast, the score of the experimental group showed a significant improvement and were notably higher than that of the control group.

Table 1. English listening and speaking test scores of the control and experimental groups before and after the training period.

	Pre-training	Post-training	P value
Control group	55.3 ± 1.2	59.4 ± 2.3	0.167
experimental group	54.6 ± 1.4	87.4 ± 1.3	0.011
P value	0.125	0.021	

The results of the questionnaire survey are shown in Table 2. The results showed that the motivation for most students to use the VR platform in the new teaching mode was to consolidate knowledge, followed by the requirements of teachers, and the rest is out of interest in the VR form; most students tended to prefer the form of VR teaching in the new teaching mode; in the new teaching mode, due to the adoption of the online discussion area of the VR platform, the anonymity of the Internet made students and teachers in a more equal position when discussing issues, and students were more relaxed, so the proportion of more active online was higher; most students believed that the VR platform in the new teaching mode can better meet learning needs, and classroom teaching and VR teaching complemented each other; most students were more satisfied with the form of the VR platform in the new teaching mode. Overall, the results of the questionnaire survey revealed that students were more inclined to the learning form of the VR platform in the VR teaching process, and most of them used the VR platform for the purpose of consolidating knowledge, and at the same time, the communication and discussion online were more active.

4 Discussion

In the era of increasingly deep globalization nowadays, English, as an international common language, has become a key tool in cross-cultural communication, academic research, and career development. Among the components of English ability, listening and speaking, as the core skills of language input and output, directly affect the communication efficiency and comprehensive language application ability of learners. However, traditional training methods of English listening and speaking are often limited to classroom teaching, audio practice, or role-playing means, which are difficult to create a real, immersive and personalized language environment, resulting in the predicament that students still "cannot understand", "cannot speak", and "cannot speak" in actual communication. The rapid development of VR technology has brought revolutionary changes to the field of education. VR technology generates a three-dimensional virtual environment through computers and combines visual, auditory, and even tactile feedback to enable users to participate in it immersively from a first-person perspective. This feature shows great potential in language teaching, especially in improving English listening comprehension and oral expression abilities, providing an unprecedented innovative path.

This paper applied VR technology to English listening and speaking training and conducted a case analysis. The LSTM algorithm used to evaluate the pronunciation of spoken English was first tested. Then, the students were divided into a control group and an experimental group. The control group adopted the traditional teaching method, while the experimental group used VR technology to assist teaching. Before and after the teaching, the English listening and speaking levels of the two groups of students were tested, and a questionnaire was used to survey the students' feelings about the VR teaching mode. The LSTM algorithm effectively rated students' English pronunciation; the students in the experimental group adopting the VR teaching mode showed more improvement in spoken English compared with the control group. The results of the questionnaire survey showed that students were more satisfied with the VR teaching mode. Reasons for the above results were analyzed. In the VR teaching mode, the introduction of VR technology into English listening and speaking training can not only construct highly simulated language scenarios, such as airports, restaurants, and conference halls, but also provide personalized content settings and interaction feedback mechanisms according to the needs of learners. For example, learners can have real-time conversations with AI-driven foreign teachers in the virtual environment to immerse in real contexts for listening discrimination and oral output; at the same time, the system can also record voice features such as pronunciation, intonation, and speech rate, and give immediate corrections and suggestions, thereby significantly improving learning efficiency and self-confidence. Moreover, the survey results of the questionnaire also showed that VR technology had the advantages of stimulating learning interest, enhancing situational memory, and promoting emotional investment. Compared with the traditional boring listening materials and

Table 2. Questionnaire survey results.

Question	Opinion	Percentage/%
What is your main motivation for learning on the VR platform?	Required by the teacher	30.2
	To consolidate knowledge	48.3
	Pure interest	21.5
Which learning form do you prefer in VR teaching?	Traditional classroom	33.2
	VR platform	66.8
Your communication status with others in the VR teaching mode	More active online	55.3
	More active offline	7.2
	Active both online and offline	37.5
Do you think the VR teaching mode can meet your learning needs?	Traditional classroom can meet the needs more	31.9
	VR platform can meet the needs more	68.1
Which part are you more satisfied with in the VR teaching mode?	Traditional classroom	29.2
	VR platform	70.8

mechanical repetitive oral practice, the immersive learning experience can attract students' attention more, enhance their sense of participation and achievement, and then form a positive learning cycle. Especially for learners who lack the environment of using foreign languages, VR technology undoubtedly opens a window to the world for them.

5 Conclusion

This article briefly introduces VR technology and the basic process of applying it to English listening and speaking training. Then, a case analysis was conducted on freshmen from the School of Foreign Languages at Hubei University of Science and Technology. The LSTM algorithm used for spoken English scoring was first tested and compared with the traditional BPNN and RNN algorithms. The students were then divided into a control group and an experimental group, with the control group using the traditional multimedia teaching mode and the experimental group using the VR platform teaching mode. The training lasted for five weeks, and the English listening and speaking proficiency was tested before and after training. The scoring error of the BPNN-based scoring algorithm was 3.6%, with a scoring time of 3.5 s; the scoring error of the RNN-based scoring algorithm was 1.3%, with a scoring time of 2.1 s; the scoring error of the LSTM-based scoring algorithm was 0.2%, with a scoring time of 1.1 s. With the increase in training time, both groups of students showed improved English speaking proficiency, but the experimental group showed a more significant improvement. The English speaking and listening scores of the two groups of students were similar before training, but after training, the experimental group showed a significant improvement.

One of the limitations of this study was that the subjects of the case analysis were only freshmen, i.e., the scope was not wide enough. Moreover, only five weeks of teaching was relatively short. Meanwhile, only LSTM was applied to assist English listening and speaking teaching in the teaching process. Therefore, the future research direction is to expand the scope of the subjects of the case analysis, increase the teaching duration, and try to use other intelligent algorithms to assist English listening and speaking teaching.

The limitation of this article lies in that only the intelligent algorithm and VR platform were used to conduct short-term training for freshmen. On the one hand, the number of students participating in the training was small. On the other hand, the duration was short, making it difficult to confirm the universality and durability of this learning method. Therefore, the future research direction is to expand the number of testees and increase the test duration. Regarding the issue of only using the LSTM algorithm in the VR teaching mode to assist in English listening and speaking teaching, in the future research direction, new intelligent algorithms will be added, not only to give pronunciation evaluation scores but also to provide the correct pronunciation. For the problem of relatively short teaching time in the case analysis, the future direction is to increase the teaching time and expand the number of students participating in the teaching mode.

Declarations

Funding

This research is one of the results of the Key Project of China Foreign Language Education Foundation (ZGWYJYJJ12Z031).

Authors' Contributions

QHY contributed to the conception of this study. QHY performed the experiments. QHY is the main contributor and writer of this manuscript.

Competing interests

The author declares that there are no competing interests.

Availability of data and materials

Data will be available on reasonable request.

References

Ding, H. and Qi, M. (2022). Situational english teaching experience and analysis using distributed 5g and vr. *Mobile Information Systems*, 2022(1):7022403. DOI: 10.1155/2022/7022403.

Duffy, C. C., Bass, G. A., Yi, W., Rouhi, A., Kaplan, L. J., and O'Sullivan, E. (2024). Teaching airway management using virtual reality: a scoping review. *Anesthesia & Analgesia*, 138(4):782–793. DOI: 10.1097/01.aoa.0001080144.95276.35.

Guo, H. and Gao, W. (2022). Metaverse-powered experiential situational english-teaching design: an emotion-based analysis method. *Frontiers in Psychology*, 13:859159. DOI: 10.3389/fpsyg.2022.859159.

Han, L. (2022). Students' daily english situational teaching based on virtual reality technology. *Mobile Information Systems*, 2022(1):1222501. DOI: 10.1155/2022/1222501.

Jiang, S., Wang, L., and Dong, Y. (2021). Application of virtual reality human-computer interaction technology based on the sensor in english teaching. *Journal of Sensors*, 2021(1):2505119. DOI: 10.1155/2021/2505119.

Li, X. and Xie, Y. (2021). Application of virtual reality technology in oral english teaching for college english majors. In *Journal of Physics: Conference Series*, volume 1820, page 012148. IOP Publishing. DOI: 10.1088/1742-6596/1820/1/012148.

Luo, X., Zhang, T., Wang, Y., and Liu, C. (2023). A study on the impact of a spherical videobased virtual reality on teaching english writing guided by experiential learning circle theory. *Journal of Educational Technology and Innovation*, 5(2). DOI: 10.61414/jeti.v5i2.125.

Ma, L. (2021). An immersive context teaching method for college english based on artificial intelligence and machine learning in virtual reality technology. *Mobile Information Systems*, 2021(1):2637439. DOI: 10.1155/2021/2637439.

Pack, A., Barrett, A., Liang, H.-N., and Monteiro, D. V. (2020). University eap students' perceptions of using a prototype virtual reality learning environment to learn writing structure. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 10(1):27–46. DOI: 10.4018/ijcallt.2020010103.

Shang, Y. (2022). Application of immersive vr virtual simulation mode in english teaching. In *2022 2nd International Conference on Social Sciences and Intelligence Management (SSIM)*, pages 9–12. IEEE. DOI: 10.1109/ssim55504.2022.10047941.

Uruthiralingam, U. and Rea, P. M. (2020). Augmented and virtual reality in anatomical education–a systematic review. *Biomedical Visualisation: Volume 6*, pages 89–101. DOI: 10.1007/978-3-030-37639-0$_5$.

Wen, J. and Fu, F. (2021). English teaching courses for students majoring in occupational health in higher vocational education based on virtual reality. In *Journal of Physics: Conference Series*, volume 1881, page 042020. IOP Publishing. DOI: 10.1088/1742-6596/1881/4/042020.

Wu, W. and Qiu, C. (2022). Deep learning analysis of english education blended teaching in virtual reality environment. *Scientific Programming*, 2022(1):8218672. DOI: 10.1155/2022/8218672.

Xie, Y., Liu, Y., Zhang, F., and Zhou, P. (2022). Virtual reality-integrated immersion-based teaching to english language learning outcome. *Frontiers in psychology*, 12:767363. DOI: 10.3389/fpsyg.2021.767363.

Zhou, Y. (2020). Retracted: Vr technology in english teaching from the perspective of knowledge visualization. *IEEE Access*. DOI: 10.1109/access.2020.3022093.