Eng Int Syst (2022) 1: 49–54

Engineering
Intelligent Systems

Oral English Training Based on Virtual Reality Technology

Dan Gao*

College of Humanities and Social Sciences, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang 163319, China

As a means of global communication, oral English ability is particularly important. In this study, virtual reality technology is applied in students' daily speaking lessons, enabling them to practice and improve their oral expression skills in a virtual environment. The quality of students' oral English was evaluated by the English pronunciation assessment system designed by the study. The results showed that students in the experimental group, whose average score was 88.18±5.01, significantly improved oral English scores through the application of virtual reality technology, while the average score of the control group was 84.64±5.6. Virtual reality technology can improve students' oral English performance and their communication skills.

Keywords: virtual reality; oral English; pronunciation score; experiment

1. INTRODUCTION

Virtual reality technology, also known as "immersive multimedia" or "computer simulated reality", is a technology that can be applied in many areas (Levak & Son, 2017). Virtual reality technology first emerged in the 20th century, and is now being applied in many fields of human activity. It includes computer and electronic information technology, as well as simulation technology (Chen, 2016). By means of computer technology, it simulates a certain environment so as to give the user the feeling of being immersed in that environment (Wiederhold, 2018; Borrego et al., 2018). At present, virtual reality technology is widely used in military training, aerospace technology, industrial engineering design, medical technology and other fields, and its role is becoming increasingly significant (Quintana et al., 2017).

As the most commonly-used language in the world, English facilitates interpersonal communication, with oral English skills playing a major role in the communication process. However, those who are learning English tend to pay more attention to input than output (Tian, 2021). This is mainly because of the lack of English context and fewer opportunities to practice pronunciation and expression. Therefore, it is important to create English language environment that gives learners the opportunity to communicate in English and improve their speaking skills.

Given the maturity of virtual reality technology, it is possible to provide a simulated environment for English learners. Thus, many English educators have gradually begun to pay attention to this technology. They incorporate virtual reality technology in the course design, so as to create an English context for learners (Lin & Lan, 2015). At present, college oral English training in China is not different from that in high school, so students have little interest in learning and lack motivation. The classroom teaching environment and teaching content lack creativity. Learners believe that the content of English textbooks is relatively boring, making it difficult to sustain students' interest. Also, the course design is too simple to meet learners' need. On the basis of the current situation, combining virtual reality technology with oral English teaching is a new way to improve the teaching and learning of English in colleges (Xiang, 2021).

This study applies virtual reality technology in oral English training. With the design and application assessment system for oral English pronunciation, the English pronunciation scores of learners can be analyzed. The role of virtual reality technology applied in oral English training is discussed, so as to promote the application of virtual reality technology in the field of English language education.

*Corresponding address: No. 2, Xinyang Road, Daqing, Heilongjiang 163319, China, Email: gdw4du@163.com

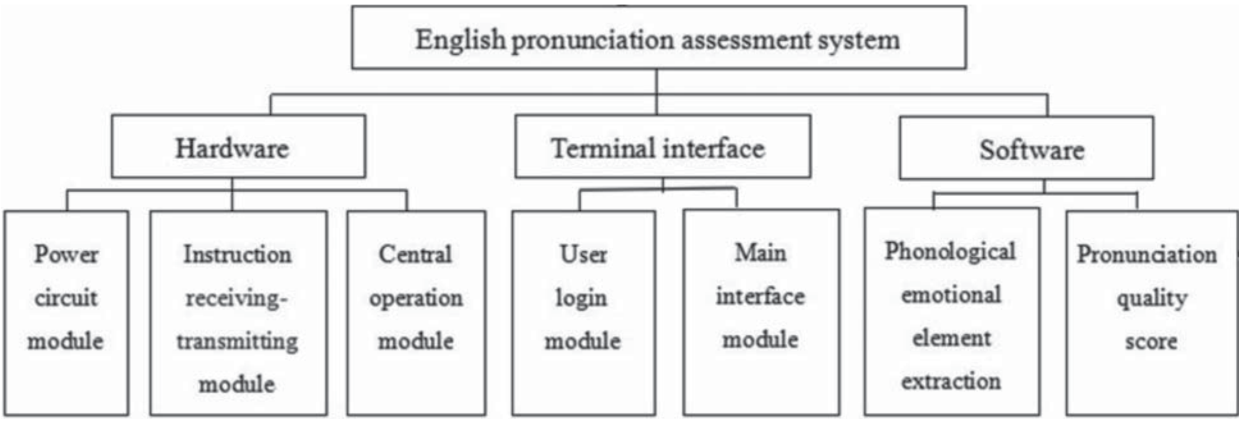

Figure 1 The overall framework of English pronunciation assessment system.

2. DESIGN OF ENGLISH PRONUNCIATION ASSESSMENT SYSTEM BASED ON VIRTUAL REALITY

The ongoing development of network technology and virtual reality technology provides technical support for English teaching. Virtual reality technology, which enables remote control and resource sharing, has become a new development direction and has good development prospects (Anderson et al., 2018). With the increase in international communication, the importance of oral English has become increasingly prominent, and the traditional manual grading of oral English pronunciation has not been able to meet the needs of the times. In order to address the current inefficiency, it is necessary to carry out an in-depth research on the assessment of oral English pronunciation.

Hence, researchers in China and abroad have studied the English pronunciation assessment system and achieved some research results. Researchers have designed a continuous interactive online English pronunciation intelligent recognition system based on MEL frequency cepstral coefficients (Polio, 2017). According to the result of speech recognition, this system selects the NOSE algorithm in the database to evaluate speech quality, check for errors and calculate the scores. This assessment system needs to calculate the scores in the database; however, the process is complex due to a large number of algorithms. Therefore, the calculation efficiency is not high. Other scholars have designed a new intelligent scoring system for oral English testing based on the Hidden Markov model (Ellis, 2017). The system consists of two modules: speech recognition and scoring process. Speech recognition applies a speech recognizer to create a Hidden Markov scoring model based on the recognition results, in order to grade English pronunciation. The hidden Markov scoring model can ensure the efficiency of grading, although the accuracy is inadequate.

Because of the problems of low efficiency and low accuracy of the above assessment system, this study designed an English assessment system based on virtual reality technology.

2.1 Overall Framework Design of The English Pronunciation Assessment System

The virtual reality-based English pronunciation assessment system is comprised of five modules: main interface, user login, power circuit, instruction receiving and transmitting, and central operation. The specific parameters of the virtual reality English pronunciation assessment system are: working temperature 0°–50°, memory 150GB, CUP7.15ghz, and disk capacity 100GB. The overall design of the assessment system is shown in Figure 1.

2.2 Hardware Design of English Pronunciation Assessment System

Each module of the English pronunciation assessment system is designed step-by-step and analyzed in detail.

2.2.1 Design of Main Interface Module

The function of the main interface module is to provide an accurate index and guidance for users. The main interface module consists of Navigation, Title, Main Content and Related Help Center. The contents of the main interface module are displayed according to the user's selection in the navigation bar. When the user enters the main interface of the system, the system's introduction and the audio database information will be displayed.

2.2.2 Design of Login Module

The basic function of the user login module is to verify the user's identity information. After entering the system, users will be prompted to enter their account and password on the interface. After the information is verified, the system will display the corresponding interface according to the user's permission. Due to the different permissions of users, the contents and information displayed on the main interface of the system need to be modified. The process of system user login is shown in Figure 2.

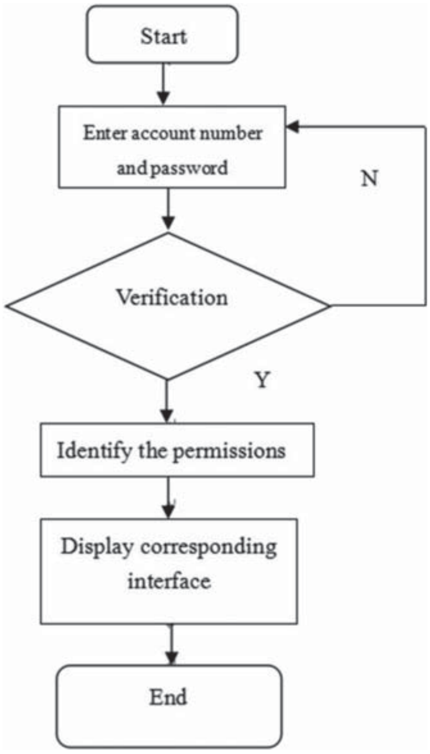

Figure 2 Chart of user login module of assessment system.

2.2.3 Design of Power Circuit Module

The power circuit module is the most important component of the virtual reality-based English pronunciation assessment system. Most of the faults and unstable factors in systems are caused by the improper design of this component. The kernel power supply of the power circuit in this system is 2V. L7806CV is used to filter the output power of the power circuit interface through C1 and C2 to stabilize the voltage of the system to 5V, and then the chip of AS1117-3.4 step-down power is used for power supply. In order to avoid external signal interference, the input and output terminals of the system are connected with their corresponding capacitors.

2.2.4 Design of Instruction Receiving- Transmitting Module

The receipt and transmission of instructions in the oral English pronunciation assessment system corresponds to the instructions on the monitoring side of the whole remote operating system. The module that receives and transmits instruction on the monitoring side generates a file containing the instruction sought by the user, while on the controlled side, this module completes the specified action of the system according to the content of the file. After the instruction is generated into a file by the central control module, the instruction receiving and transmitting module is activated to extract and assess the contents in the file, and corresponding functions are performed according to the instruction.

2.2.5 Design of Central Operation Module

The main function of the central operation module of the virtual reality-based English pronunciation assessment system is to analyze and process the received instructions. The central operation module is the main part of the controlled terminal program of the system. It receives various instructions through the listening port and processes them. After processing, it receives the next instruction from the listening port of the system, and the process is repeated until the system is shut down. According to the design and analysis, the main interface module in the virtual reality English pronunciation assessment system can act as a guide to help users quickly find the required business. The user login module can verify the user's information and ensure the user's security. The power circuit module can filter the external signal interference, reduce the failure rate, and ensure the smooth experience of the system's users, The instruction receiving and transmitting module is responsible for executing the received instruction. The central operation module first analyzes the instructions, then processes them until it completes the assessment of user's English pronunciation and obtains the scores.

3. SOFTWARE DESIGN OF ENGLISH PRONUNCIATION ASSESSMENT SYSTEM

On the basis of the hardware module design in the English pronunciation assessment system, the PID algorithm is adopted to extract the emotional elements of English pronunciation in a hierarchical way. Then the quantitative recursive analysis method is used to evaluate the quality of the English pronunciation. After that, the final score is obtained.

3.1 Extraction of Phonetic Emotional Elements

The current oral English pronunciation assessment systems consider only the tone, intonation and rhythm of speech. Unless the role of speech emotion is considered, the assessment of English pronunciation is not comprehensive. In order to solve this problem, the PID algorithm is proposed to extract emotional elements in oral English in a hierarchical manner. Taking the imbalance of evaluation data into full consideration, various factors affecting oral English pronunciation are extracted in the virtual reality-based pronunciation assessment system. Since the traditional system extracts only the conventional indexes of tone, intonation and rhythm, the PID algorithm is adopted to extract the phonetic emotional elements of oral English so as to accurately score the pronunciation quality. PID is the most commonly-used algorithm for remote operation (Matsuda, 2017). It is assumed that the actual output value of the virtual reality-based English pronunciation assessment system is $C(t)$ and the fixed value is $R(t)$. The calculation formula for the operating deviation of the system is as follow:

$$e(t) = c(t) - r(t) \tag{1}$$

3.2 Realization of Evaluating Oral English Pronunciation Quality

To obtain results for phonetic emotion elements, a quantitative recursive analysis method is used to comprehensively evaluate the quality of oral English pronunciation, and finally the scores are obtained. In the optional evaluating methods, the accent difference is calculated according to the accent of each phonetic unit in the speech to be evaluated and the standard answer. It can be calculated as follows:

$$\begin{aligned} diff = \sum_{i=1}^{n} \Bigg\{ & \left(\frac{left_{std}[i])}{Len_{std}} - \frac{left_{test}[i])}{Len_{test}} \right) \\ & + \left(\frac{right_{std}[i])}{Len_{std}} - \frac{right_{test}[i])}{Len_{test}} \right) \Bigg\} \end{aligned} \tag{2}$$

In this formula, *diff* signifies accent difference; n is the number of accent units in oral English; Len_{STD} is the frame length of speech material in the standard answer; $left_{STD}[i]$ is the starting frame position of the ith accent unit in the speech material of standard answer; $right_{STD}[i]$ is the terminal frame position of the ith accent unit in the speech material of standard answer; Len_{test} is the frame length of the speech material to be evaluated; $left_{test}[i]$ is the starting frame position of the ith accent unit in the speech material to be evaluated; The $right_{test}[i]$ is the terminal frame position of the ith accent unit in the speech material to be evaluated. In other words, the short-time energy characteristic curve is used to obtain the accent difference between the speech materials to be scored and the ones in the standard answer. According to the accent difference obtained the score is conducted, which greatly reduces the calculation amount, improves the scoring efficiency, and achieves the assessment of oral English pronunciation quality.

The scoring process used in the method above is implemented through the instruction of the computer program and the corresponding hardware facilities. These programs can be stored in a computer or in a readable device such as a disc, a compact disc, a flash disk, a read-only memory (ROM) or a random access memory (RAM).

4. EXPERIMENTAL RESEARCH ON ORAL ENGLISH TRAINING UNDER VIRTUAL REALITY ENVIRONMENT

4.1 Test Subjects

In this study, the test subjects were 90 undergraduates studying English at a university in Heilongjiang province. The subjects were between 20 and 21 years old. Before the end of the second year, they had passed TEM-4 (Test for English Majors-4), and they were all in the first semester of their junior year when tested. Compared with other English learners, English majors invest more time in English learning and are likely to be engaged in English learning in the long term. Therefore, English majors were chosen as the test subjects, as it was considered more productive to investigate the differences in their level of oral English competency.

4.2 Experimental Process

For the research task, the experimental group was given an eight-week oral English training program. Students in the experimental group learned with the assistance of a virtual reality game (Scendlife) (Cheok & Zhang, 2019; Parr & Friston, 2017). The students in the control group learned exactly the same content as the students in the experimental group, but they used traditional multimedia technology, which can help to display the content of teaching materials better. Both groups learned autonomously. Teachers did not interfere with the learning process. At the end of the experiment, the subjects were tested for their proficiency in spoken English. The test included English reading materials in textbooks and the presentation of specific topics.

4.3 Experimental Data Processing Methods

In this study, before the oral English test, subjects were randomly divided into two groups. A double-population T-test was used to analyze the data. S_2^1 and S_2^2 were variances of the two samples; n_1 and n_2 are the sizes of the two samples.

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{3}$$

Table 1 English scores of the experimental group before and after experiment.

	Before	After	T	P
Experimental group	81.54±5.82	88.18±5.01	−4.63	0.00
Control group	81.65±4.24	84.60±5.60	2.46	0.02

Table 2 Comparison of experimental group and control group before and after the experiment.

	Pre-test scores			Post-test scores		
	100–90	79–89	79–70	100–90	79–89	79–70
Experimental group (44)	11.37%	50.00%	38.63%	36.36%	59.09%	4.55%
Control group (46)	13.04%	50.00%	36.96%	17.39%	54.35%	28.26%

Table 3 Comparison of oral English scores for experimental group and control group.

	Experimental group	Control group	T	P
Between the experimental and control groups	88.18±5.01	84.64±5.6	2.153	0.038

T test (two-side test, test level: =0.05) was used to statistically process the TEM-4 scores of the two test groups, $t = 0.150$, $P = 0.968 > 0.05$, indicating that X fell within the confidence interval, thus indicating that there was no significant difference in the average TEM-4 scores of the two test groups. Therefore, it can be assumed that the overall English level of the two test groups is equivalent.

4.4 Results and Analysis

4.4.1 The Enhancement Effect of Virtual Reality Technology Assisting Oral English Training

Analysis of Oral English Ability in Experimental Group and Control Group After Experiment SPSS26.0 was used for statistical analysis. The 90 subjects were divided into two groups: 44 in the experimental group and 46 in the control group. In the experimental group, the percentage of students scoring 100–90 points increased from 11.37% to 36.36% (the number of students increased from 5 to 16). The percentage of students with 89–80 points rose from 50.00% to 59.09% (the number rose from 22 to 26). The percentage of students with 79–70 points dropped from 38.63% to 4.55% (the general number dropped from 17 to 2). In the control group, the percentage of students with 100–90 points increased from 13.04% to 17.39% (the number increased from 6 to 8). The percentage of students with 89–80 points rose from 50.00% to 54.35% (the number rose from 23 to 25). The percentage of students with 79–70 points dropped from 36.96% to 6.52% (the general number dropped from 17 to 3). Using SPSS26.0 paired sample T test, it can be seen from Table 1 that the oral English pre- and post-test results for the experimental group P=0.00, which is less than 0.05, so there is a significant difference. In addition, the mean score for the oral English post-test was 88.00, which was higher than that for the pre-test (81.54). Therefore, the experimental group's score for the oral English post-test was significantly better than that for the pre-test. The pre- and-post test results of the control group P=0.02 is less than 0.05, showing an obvious difference. Moreover, the mean score for the pre-test is higher than that for the post-test. Evidently, after a semester of training, the oral English scores of both the experimental group and control group increased. It can be seen from table 1 that, while both the oral English levels of the experimental group and control group improved, the amount of improvement in the scores of the experimental group is much higher than that of the control group. This indicates that virtual reality technology can be used successfully to teach oral English; moreover, this technology can significantly increase the students' level of proficiency, more so than traditional methods.

Analysis of Oral English Ability Between Experimental Group and Control Group After Experiment After the VR assisted training experiment, the experimental group and the control group were given a post-test in the eighth week, and the subjects were tested for their oral English ability with the English pronunciation assessment system. As can be seen from Table 2, after virtual reality assisted learning, the increase rate of oral English scores of the experimental group was significantly greater than that in the control group.

In the experimental group, those with 100–90 points accounted for 36.36% of the total number, while those with 100–90 points accounted for 17.39% of the total number in the control group. In the experimental group, 59.09% of students scored 89–80 points, while in the control group, 54.35% scored 89–80. The number with 79–70 points in experimental group accounted for 4.55% of the total and the number with 79–70 points in control group accounted for 28.26% of the total number. The following form was obtained with SPSS26.0, by T test of the independent sample, $P = 0.038 < 0.05$ (table 3), so it is significantly different. The mean value of experimental group is higher than that of the control group, thus oral English scores of the experimental group is better than that of the control group.

It can be concluded that virtual reality technology can improve the oral skills of English language learners. There are several reasons for this: virtual reality technology overcomes the limitations of time and space, creates a good English language environment for learners, integrates the learning contents into the created situation, and offers the learner a highly realistic experience. The "immersive scene" can engage the learners and encourage them to actively participate in English learning. Moreover, learners can shift from passive learning to active learning according to their individual needs. At the same time, this kind of virtual situation learning can reduce the learning pressure and create a relaxed learning atmosphere which is conducive to better learning.

5. CONCLUSION

In this study, comparative experiments were conducted to determine whether the use of virtual technology as a learning tool has an impact on students' oral language skills. From the test results, it can be concluded that virtual reality technology can significantly improve English language learning and can significantly improve the outcomes of oral English teaching. Virtual reality technology, by simulating the real situation, gives learners an immersive feeling and stimulates their interest in learning, enabling them to shift from passive learning to active learning. To date, there has been little research on the use of virtual technology for oral English teaching, possibly because it is a major challenge for teachers. It requires teachers not only to constantly learn new ways of teaching, but also to master the advanced teaching technology and transform them from traditional teachers to active researchers. Moreover, the use of this technology as a teaching tool in the classroom might be of no benefit to students who lack self-discipline. As an emerging technology, it is necessary to further explore the application of virtual reality technology in oral English teaching. It is anticipated that this study will be a valuable reference for researchers interested in exploring the application of virtual technology in the classroom as a means of improving students' oral English ability.

REFERENCES

1. Anderson, J. A. E., Chung-Fat-Yim, A., Bellana, B., Luk, G. & Bialystok, E. (2018). Language and Cognitive Control Networks in Bilinguals and Monolinguals. *Neuropsychologia*, (6), 1–47.
2. Borrego, A., Latorre, J., Alcaniz, M. & Llorens, R. (2018). Comparison of Oculus Rift and HTC Vive: Feasibility for Virtual Reality-Based Exploration, Navigation, Exergaming, and Rehabilitation. *Games for Health Journal, 7*(3), 4–14.
3. Chen, J. C. (2016). The crossroads of English language learners,task-based instruction, and 3D multi-user virtual learning in Second Life. *Computers & Education*, (102), 152–171.
4. Cheok, A. D. & Zhang, E. Y. (2019). Kissenger: Transmitting Kiss Through the Internet. *Reviews on Biomarker Studies in Psychiatric and Neurodegenerative Disorders.*
5. Ellis, R. (2017). The Production-Oriented Approach: Moving Forward. *Chinese Journal of Applied Linguistics*, (4), 454–458
6. Levak, N. & Son, J. (2017). Facilitating second language learners' listening comprehension with Second Life and Skype. *ReCALL*, *29*(2), 200–218.
7. Lin, T. & Lan, Y. J. (2015). Language Learning in Virtual Reality Environments: Past, Present, and Future. *Journal of Educational Technology & Society*, *18*(4), 486–497.
8. Matsuda, P. K. (2017). Some Thoughts on the Production Oriented Approach. *Chinese Journal of Applied Linguistics*, (4), 468–469.
9. Parr, T. & Friston, K. J. (2017). The active construction of the visual world. *Neuropsychologia*, (8), 1–22.
10. Polio, C. (2017). Reflections on the Production-Oriented Approach vis-á-vis Pre-service Teachers. *Chinese Journal of Applied Linguistics*, (4), 464–467.
11. Quintana, M. G. B., Sagredo, A. V. & Lytras, M. D. (2017). Pre-service teachers' skills and perceptions about the use of virtual learning environments to improve teaching and learning. *Behaviour & Information Technology*, (8), 1–14
12. Tian, M. (2021). A study of English listenting and speaking teaching mode for college students: using multimedia network technology. *Engineering Intelligent Systems*, 29(1), 5–10.
13. Wiederhold, B. K. (2018). When Second Life becomes real life: the evolution of self-presentation. *Cyberpsychology Behavior & Social Networking*, *21*(1), 1.
14. Xiang, L. Y. (2021). Lifelong learning:?a study of college English autonomous learning. *Engineering Intelligent Systems*, 29(2), 103–108.