Modern Management based on Big Data III
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220086

Applications of GMM-HMM Acoustic Model in the Immersive Foreign Language Learning

Jianghui LIU[1], Zitong WENG and Yiming ZHONG
Guangdong University of Foreign Studies, Guangzhou, China.

Abstract. This research presents an immersive oral English teaching mode by combining optimized GMM-HMM、VR technology with immersive learning theory, which enables learners to learn English in a real context with .This model can cultivate learners' cultural awareness and enhance their output ability in an effective way, making them achieve self-innovative development in the process of independent English learning. Informed by the current development of VR technology and English teaching through literature research, this study centers on the designs of immersive VR teaching models. It presents some practical experience by undertaking comparative research which was implemented based on immersive VR teaching and multimedia teaching. This study aimed to facilitate the trend of combining information technology and education.

Keywords. Optimized GMM-HMM acoustic model, immersion, VR, foreign language learning

1.Introduction

VR (Virtual Reality) information technology, also known as Virtual Reality simulation information technology, is a combination of computer science, sensor technology and performance Technology. In the virtual space formed by computer science, users can use head-mounted displays for real-time control, move freely in this space, and interact with the virtual environment with the help of multi-sensory channels such as sight, hearing and touching [1]. In recent years, with the rapid development of science and technology, virtual reality technology has been applied to many fields. Among them, the combination of VR technology and education is now a development direction with broad prospects [2-3].

In college, English teaching usually focuses on students' language application ability, but the expressions can only be trained and improved in the real environment of life. However, at present, many students are generally accustomed to the exam-oriented education mode in high school, and their oral output skills are relatively weak. In universities, the interaction between teachers and students in large classes is often very limited, and students have few opportunities to speak and practice English. The

[1] Corresponding Author, Liu Jianghui Guangdong University of Foreign Studies, Guangzhou, China; Email: 247031690@qq.com
This study was financially supported by the Undergraduate Innovation Training Project of Guangdong University of Foreign Studies in 2022.

phenomenon of "mute English" in class will further reduce students' interest in English learning. It is difficult for students to speak English at the beginning and improve their practical communicative ability.

In view of this situation, it is necessary to combine virtual reality technology with oral English teaching and give full play to the advantages of VR's "3I" characteristics, namely Imagination, Interaction and Immersion [4]. Teachers can teach by using virtual reality technology to construct real situations where students can get the feeling of "being on the scene", so as to strengthen language training and teaching of expression.

2. Research background

Immersive classroom teaching is a teaching method that emphasizes the use of the second language as the teaching language, that is, learners are "immersed" in the target language environment [5]. It originated in the 1960s in The United Kingdom. At that time, English and French were the official languages of the United Kingdom. However, due to the sharp increase of the immigrant population in the United Kingdom, the number and demand of learning and using English gradually increased, while French was the opposite. Based on this, the federal government took measures to improve the state of French education, which also led to the birth of immersion. After the success of immersion teaching in the UK, it has been used for reference by many countries and regions [6], which has greatly promoted the development of bilingual education.

In 1996, China launched a project on immersion English teaching, comprehensively reforming the original English teaching methods. On the basis of the previous teaching mode of English grammar, colleges and universities provide more opportunities for English output, hoping to improve students' English application ability and improve the overall English level. Immersion English teaching mode requires students to completely "immerse" themselves in the learning environment of the target language, so as to promote students to develop English thinking habits in a relatively short time, so as to achieve flexible use of English [7].

To achieve immersion learning, it is best to let students really enter a situation close to real life to feel and experience [8]. However, due to practical problems such as time, cost, distance and safety, it is difficult to reproduce the whole scene in the actual teaching process. The application of VR technology creates possibilities for immersive English teaching [9]. Immersion English enables learners to closely relate theory to reality after learning it. In the virtual reality learning environment, learners no longer only memorize and recite monotonously, but on this basis enhance the full application of theoretical knowledge in the real situation. The precise simulation and real-time interaction of VR skills can affect the quality of classroom teaching [10][11].

3.VR Immersive Oral English Learning System based on optimized GMM-HMM Acoustic Model

The generation process of the optimization of GMM-HMM acoustic model is shown in Fig. 1.

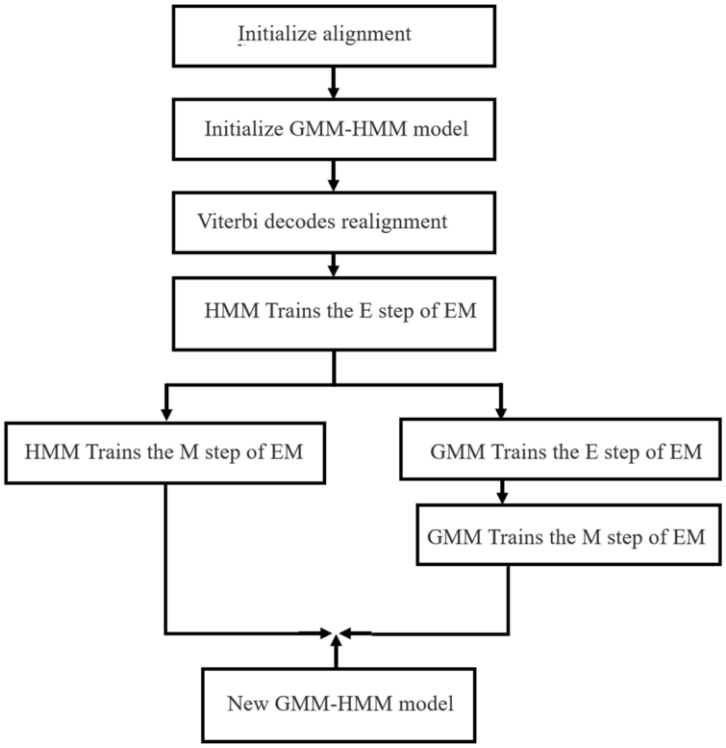

Fig. 1 Optimization of GMM-HMM acoustic model

The system will use the Viterbi training method (based on EM algorithm), which needs to explicitly input the state of each frame, and use marked data to update the parameters of GMM. Under this algorithm, it can run faster, but the performance of the model has no obvious loss.

- Model initialization. To carry out the training of HMM model, model parameter is taken to make $P\ (O \mid \lambda)$ local maximum values. Hence, choose a good initial model, to make the local maximum of parameters and global great closer, eventually improve the overall effect of the model.
- Then explicitly input the state of each frame and obtain the state number of each frame as the training label. Firstly, the state sequence of the trained speech can be obtained, and a state graph can be obtained, which gives the state sequence of the trained speech, and then the alignment is completed through the decoder.
- Finally, conduct model training using three-tone submodel training. When designing modeling units, we should not only consider the central phoneme itself, but also consider the context phoneme where the phoneme is located, which is the context dependent acoustic model, namely the three-tone submodel. In actual scenes, there may be coarticulation, and the actual pronunciation of phonemes may be affected by neighboring and similar phonemes, or may be changed due to different positions in sentences. Therefore, in order to improve the performance of speech recognition system, a more realistic three-tone submodel is used for modeling.

ER algorithm:

Input: observation data x= (x1,x2... , xm), the joint distribution $p(x,z|\theta)$, conditions of step-by-step $p(z|x,\theta)$, the largest number of iterations J

- Randomly initializes the initial value of the model parameter θ^0
- For j from 1 to J, the EM algorithm iteration starts:

a) E step: calculate the conditional probability expectation of the joint fractional steps:

$$Q_i(z^{(i)})=P(z^{(i)}|x^{(i)},\theta^j) \quad (1)$$

$$L(\theta,\theta^j)=\sum_{i=1}^{m}\sum_{z(i)}Q_i(z^{(i)})logP(x^{(i)},z^{(i)}|\theta) \quad (2)$$

b) M step: maximize $L(\theta,\theta^j)$, obtain θ^{j+1}:

$$\theta^{j+1}=arg\max_{\theta}L(\theta,\theta^j) \quad (3)$$

c) If θ^{j+1} converges, the algorithm ends, otherwise go back to step a) for step E iteration.

Output: model parameters θ

4.Experiment and Results

4.1 Design

Two different teaching modes are defined as independent variables, namely immersive VR teaching mode and traditional multimedia teaching mode. The differences and similarities between immersive VR teaching mode and traditional multimedia teaching mode are investigated.

The primary objective of this study was to determine whether immersive VR English teaching (independent variable) would improve learners' oral English. The hypothesis is that compared with the control group, oral performance will be significantly improved due to the application of immersive VR experience in the experimental group.

Taking into account these facts, the following primary research hypotheses were formulated:

Hypothesis 1. Learners who had immersive VR experience during the English course had higher spoken English fluency than those who did not.

Hypothesis 2. The oral English accuracy of learners who had immersive VR experience during the English course was higher than that of learners who did not.

A secondary goal of the study was to track participants' emotional state after the intervention. The secondary dependent variables were measured by two scales - the learning motivation questionnaire and the English Learning Anxiety Scale.

Hypothesis 3. The learners who got immersive VR experience during the English course were more motivated to learn English than those who did not.
Hypothesis 4. Learners who had immersive VR experience during the English course had lower anxiety in English learning than those who did not.

4.2 Participants

A total of 30 sophomore undergraduates from a university in Guangdong province participated in the study. Their average age was 20. One class was assigned to the experimental group (Class A) and the other to the control group (Class B). Each of them consists of 15 students. None of the participants had prior experience using immersive VR in English classes. For computer-assisted language learning, undergraduates often read audio files or watch video files played by their teachers as they practice their spoken English skills.

4.3 Materials

Take Unit 1 “Getting to Places” as an example. This Unit mainly introduces the transport system and landmark location of London, involving historical background knowledge, spatial direction and other related vocabulary and expressions. In traditional Chinese teaching, teachers only introduce various landscapes to students with the help of pictures in books and PPT, and teach relevant background and professional knowledge. However, students cannot truly feel exotic scenery and humanistic scenery, and it is difficult to form a deep understanding of “Double-Decker”, “Minicab” and other British characteristic transportation vehicles, let alone form a deep understanding of a more abstract concept of orientation in a completely unfamiliar city.

4.4 Instruments

After the students in the experimental group and the control group completed the weekly course learning, they were required to record a piece of audio composition according to the task assigned by the teacher (related to the theme of this unit), and the teacher scored according to the scoring guide.

4.5 Results

In general, through the comparison of test scores and scale results, the following findings are obtained: first, immersive VR teaching mode has a positive impact on students' oral performance, specifically in terms of improved language fluency, comprehension and maturity. Therefore, it can be inferred that immersive VR can improve the effect of English teaching; Second, immersive VR teaching mode can improve students' motivation for English learning. It can be inferred that immersive VR brings students multi-sensory stimulation and helps stimulate their interest in English learning. Thirdly, immersive VR teaching mode is helpful to relieve English learning anxiety of learners. Therefore, it can be inferred that students will feel more confident and safer when facing virtual characters, and thus use English more actively and boldly.

5.Conclusion

To sum up, immersive VR technology brings great possibilities for English teaching. The integration of virtual reality technology and English has created a new English teaching environment, which is helpful to solve the problem of context that has long puzzled English learners in China. The learning modes proposed in this paper can enable learners to experience and actively participate in real situational activities in a virtual way, further develop their discourse and logical thinking in real situational activities by using the professional English knowledge they have learned, practice boldly, and experience success in them. But at the same time, VR technology in the application of classroom teaching, still face many practical problems. VR technology to create a new model of classroom teaching is still only a prototype, there is still a large gap between imagination and improvement. In general. The application of virtual reality technology, especially immersion VR, in higher education is still not fully mature. Hardware facilities, application software resources, application environment teaching methods and evaluation and other application fields need to be further studied and discussed.

References

[1] Gardner, R.C., & Lambert, W.E. (1959). Motivational variables in second-language acquisition. Canadian journal of psychology, 13, 26-72.
[2] González C R, & Martín-Gutiérrez J, & Domínguez M G, et al. (2013) Improving spatial skills: An orienteering experience in real and virtual environments with first year engineering students. Procedia Computer Science, 25: 428-435.
[3] Aydogan H, Ata R, Ozen S, et al. (2014) A study of education on power transformers in a virtual world. Procedia -Social and Behavioral Sciences, 116: 3952-3956.
[4] Burdea G, & Coiffet P. (2003) Virtual reality technology, second edition. New York : John Wiley & Sons:3-4.
[5] Ke Ren. (2021). An Analysis of the Application of Immersion Teaching Method in Japanese Teaching. Journal of International Education and Development (9).
[6] PhD T. J. Ó Ceallaigh. (2016). Second Language Acquisition and Form-Focused Instruction in Immersion: Teaching for Learning. World Journal of Educational Research (2).
[7] Richardson, J. C., & Swan, K. (2003). Examining social presence in online courses in relation to students' perceived learning and satisfaction. Journal of Asynchronous Learning Networks, 7(1): 68-88.
[8] Grassini S, Laumann K, Rasmussen Skogstad M. (2020). The Use of Virtual Reality Alone Does Not Promote Training Performance (but Sense of Presence Does). Front Psychol, (11) : 1–17.
[9] Wang C, Lan Y J, Tseng W T, et al. (2019) On the effects of 3D virtual worlds in language learning–a meta-analysis. Computer Assisted Language Learning: 1-25.
[10] Jorge Bacca-Acosta, Julian Tejada, Ramon Fabregat, Kinshuk, Juan Guevara. (2021) Scaffolding in immersive virtual reality environments for learning English: an eye tracking study, Educational Technology Research and Development, 2021.
[11] Lan, Y. J. 2020. Immersion, Interaction and Experience-oriented Learning: Bringing Virtual Reality into FL Learning. Language Learning & Technology(1) :1-15.