J. COMBIN. MATH. COMBIN. COMPUT. 124 (2025) 333–349

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A virtual simulation-based perspective on intercultural communication in language learning

Li Chen[1,✉]

[1] *Department of Chinese Language and Literature, Pingdingshan Vocational and Technical College, Pingdingshan, Henan, 467000, China*

ABSTRACT

The application of virtual reality (VR) technology in teaching is increasingly widespread. This study leverages VR to create cross-cultural teaching contexts and develop speech recognition models for language learning. An ecological model of language learning based on VR is constructed, and a cross-cultural contextual VR system is implemented and introduced into language education. Testing reveals that the system achieves a speech recognition efficiency of 99.7% and a correctness rate of 99.5%. Moreover, a comparison of pre- and post-test data between experimental and control groups shows that the experimental group significantly outperformed the control group in English proficiency ($p < 0.05$). Overall, the cross-cultural contextual VR system demonstrates a significant positive impact on language learning outcomes.

Keywords: virtual reality, context creation, intercultural communication, language learning

1. Introduction

Language has been shown to be related to the dynamics, movements, and perceptual experiences of our bodies. That is, the cognitive processes underlying the types of language are for us directly influenced by our bodily reality and its relationship to our bodies [12, 7]. Even the way we construct metaphors in language is directly aligned with body-centered associations.

Often proposed as a platform for providing embodied learning, which allows students to engage in activities closely related to their bodily movements in new environments, virtual reality is emerging as a promising platform for kinesthetic language learning, i.e., learning through bodily activity, and is deeply rooted in the relationship between our minds and our bodies [16, 5].

✉ Corresponding author.
E-mail address: 19937516797@163.com (L. Chen).

Received 12 February 2024; Accepted 02 January 2025; Published Online 17 March 2025.

DOI: 10.61091/jcmcc124-23

On the other hand, intercultural communication encompasses not only verbal communication, but also non-verbal modes of communication, such as body language and the use of time and space, which are significantly culturally differentiated. Intercultural communication is indeed a complex competence that transcends the visible dimension of culture and embraces many elements of the invisible dimension of culture, i.e., those factors that are based on beliefs and values specific to a given social group and thus govern its behavior [9, 6]. Intercultural communicative competence is therefore an area of research that is becoming increasingly important in the increasingly multicultural communities in which we live - especially in teaching/learning environments.

Virtual Reality Language Learning System, which strengthens the connection between word-action pairs by using artificial intelligence gestures to recognize the actions displayed by students in a virtual reality environment and presenting the corresponding names of the actions performed in the target language. Virtual reality technology can influence language learning through the use of kinesthetic elements. Parmaxi [15] reviewed the research literature on virtual reality-based language teaching and learning.Topics included the performance of virtual reality technology in language learning, as well as directions for future research. It was concluded that virtual reality technology is an effective tool in language teaching and learning, but there are still deficiencies in the configuration of the technology and the corresponding instructional design. Legault et al. [13] conducted teaching experiments to confirm the importance of language learning environments and noted that students' second language scores gained significant progress in an immersive language teaching classroom based on virtual reality technology. Cheng et al. [8] designed a virtual reality game for Japanese language learning and found that this language learning model was beneficial for enhancing participants' understanding of language and culture, as well as helping students to learn a second language. Alfadil [3] Based on a sample language comparison experiment, it reveals that the Language House game with virtual reality technology as the core logic can promote students' vocabulary acquisition in the process of language learning, and effectively assist language teachers to enhance the effect of language teaching. Tai et al. [18] examined how VR technology affects students' English vocabulary learning effects, and therefore held a teaching experiment to analyze the relevant research, which showed that the virtual reality technology-enriched language classroom effectively enhances the interactivity of the language teaching classroom and provides rich scenarios for vocabulary learning, which effectively stimulates the students' motivation for language learning. Scholars have affirmed the positive role played by virtual reality technology in language teaching, so the future research direction of virtual reality technology in language teaching practice should be the design of teaching methods, teaching content selection and corresponding teaching management adapted to virtual reality technology.

Intercultural communication competence is a complex competence that transcends the visible dimensions of culture, and cultural mediation places intercultural competence at the center of language teaching. House et al. [11] based on a cross-cultural pragmatics perspective, assembled surveys, interviews, and discourse completion tests to design a methodology for the study of second language learning, which supports readers to rely on the evidence of pragmatics to help readers understand and learning a second foreign language, and discussed in detail with some practical cases . Alqudah et al. [4] explored the association between user happiness, learning effectiveness, etc. and cross-cultural semantic differences in the use of language learning online platforms, and based on the results of the analysis, it was learned that the cultural association has a positive significance on the improvement of user learning effectiveness and happiness. Roberto [17] describes a technology-enhanced language learning strategy that encompasses three cross-cultural languages, i.e., natural language (which supports cross-cultural language switching), visual language, and artificial language, and provides an in-depth analysis of the underlying logic of the technology-enhanced language learning strategy and the connections with the three cross-cultural languages. Ahtif [2] explored the issues related to language communication in cross-cultural contexts, emphasized the important role of language in the process of cross-cultural communication, and put forward targeted and effective suggestions to promote effective communication among people with cross-cultural backgrounds. Scholars have clarified the significance of intercultural competence for language learning, but there is still much room for reflection and innovative approaches to intercultural competence development.

Based on the importance of cross-cultural communication in language learning, the article adopts virtual reality technology to simulate cross-cultural communication scenes in language learning, create cross-cultural communication situations, construct cross-cultural communication speech recognition models and carry out contextual teaching. The ecological model of language learning based on virtual reality is constructed, and on this basis, the cross-cultural contextual teaching/learning VR system for language learning is further designed and developed. Subsequently, the speech recognition performance of the language learning system is tested, and then the learning interactions in the cross-cultural communication context in the language learning system are analyzed. The effectiveness of the cross-cultural context teaching/learning system proposed in this paper is verified by implementing teaching experiments and comparing the pre- and post-test language levels of the experimental and control groups.

2. Virtual reality-based intercultural communication situation construction

2.1. Virtual reality

Virtual Reality (VR) is a cutting-edge technology in the field of computing capable of generating a virtual environment that simulates the real world [1]. The use of computer simulation technology and a variety of multimedia technologies to build a realistic virtual scene, the user in the computer software and a variety of sensors support, can be formed through the human visual, auditory, tactile and other senses of human interaction with the machine, in this immersive environment for a variety of interactive experience. Among them, the sensors in virtual simulation technology include data gloves, three-dimensional spatial interaction balls, helmet displays, operating handles and other sensing devices. Virtual reality technology provides a new interactive mode of operation for people and computers, which is mainly characterized by:

2.1.1. Immersion. Immersiveness, also known as presence, refers to the degree of realism that the user feels as the protagonist exists in the virtual environment, and is considered to be the performance scale of the VR system. Virtual reality technology based on the physiological and psychological characteristics of human vision, hearing, by the computer to produce realistic three-dimensional stereo images, the user in the data gloves, helmet display, operating handle and other sensing equipment, you can put yourself in the virtual environment, resulting in an immersive feeling.

2.1.2. Interactivity. In virtual reality, the user uses the sensing equipment to touch or operate the objects in the virtual environment, the computer will update the three-dimensional view in time, and present the new visual effects to the user, which is highly interactive.

2.1.3. Conceptualization. Designers according to certain goals, through the imagination and the creation of virtual environments. Users can enter the virtual space, rely on their own perception and cognition of the state and process of the system operation for logical thinking and associative creation, in order to obtain new knowledge.

2.2. Intercultural communication in language learning

The goal of language learning is not only the learning and mastering of systematic basic theoretical knowledge and professional knowledge, but also focusing on or providing the intellectual background and methodology for research in specialized fields. Language learning is to realize intercultural communication.

Cultural communication has been integrated into our value system, our social behavior and demeanor. Thus, the composition of culture is part of building communicative competence. Communicative competence is conceived as a combination of knowledge belonging to the target culture, and it can anticipate misunderstandings, avoid sudden interruptions in linguistic communication, and explain situations of conflict. Nowadays, language learning should take more account of the cultural dimension, since any contact between two different language systems will inevitably lead to the interaction of two very different cultural systems.

Learning a language also involves learning the culture behind that language, and is a continuous process of cultural reconstruction for the beginner. Through exposure to another language and culture, one acquires a faith and is able to discover the identity of that culture.

2.3. Situational approach

Contextual teaching is categorized into virtual contextual teaching and real contextual teaching [10, 14]. The context studied in this paper is the virtual context of language learning. Virtual context usually refers to the use of information technology to create and simulate the "real" scene, which can create a learning environment similar to the real situation for learners. In the process of teaching, sometimes the original environment around the students can not meet the needs of teaching, so it is necessary to appropriately virtualize part of the environment. In virtual context teaching, we should guide students to truly integrate into the virtual teaching situation, stimulate students' interest in geography learning and imagination, so that students from passive acceptance of knowledge to active exploration of knowledge, the development of innovative consciousness, and then achieve better teaching results. The virtual context facilitates students' efficient learning and knowledge understanding by simulating real scenes.

2.4. Intercultural communication context creation

2.4.1. Creation of concepts. Virtual simulation learning contexts are created on a virtual simulation environment platform. It mainly has the following characteristics:

(1) *Support for multi-user participation in the virtual environment.* In the virtual simulation cross-cultural communication situation, users participating in cross-cultural communication can log in the same location at the same time. So that the virtual workspace in which the users are located is the same, but the user's authority in the virtual workspace will be different.

Multi-user participation in a virtual simulation intercultural communication context allows users to participate together in a virtual simulation space and interact in the space. This virtual space is similar to the real space, and with a shared virtual workspace, users can participate and collaborate in the virtual simulation cross-cultural communication context.

(2) *Operate the avatar for interaction.* The virtual simulation cross-cultural communication environment can provide users with a virtual avatar, which is a virtual reproduction of the operator's entity, and the operator can get a sense of reality similar to real life by operating the virtual avatar.

The virtual simulation of cross-cultural communication context of this incarnation interaction provides a virtual way to establish a direct similarity to the real-life experience of the possibility, we are in the virtual simulation of cross-cultural communication context of foreign language learning process, the learner can get similar to the real-life experience, the incarnation of interaction is an important part of the virtual simulation of cross-cultural communication context of experience acquisition, is an important part of the virtual simulation of cross-cultural communication context. It is an important part of the virtual simulation intercultural communication context.

(3) *Support real-time interaction.* All users logging into the virtual simulation intercultural communication context need to abide by the time of the virtual simulation intercultural communication context. Virtual simulation of intercultural communication situations have their own independent time, this time is fair to each participating user, and is independent from the real time.

(4) *Compatible with various types of media interaction.* In the virtual simulation cross-cultural communication context, there are various ways of interaction between users, and the media that can be integrated can also be said to be fully compatible. In the case of online distance learning, the diversity of interaction modes is a prerequisite for the development of diverse learning activities. Analyzing the advantages and disadvantages of various interaction modes and utilizing them reasonably will make the learning process smoother, which has a significant role in the virtual simulation of cross-cultural communication situations.

(5) *Full-angle virtual simulation operation.* Users can operate virtual objects in a natural way in virtual reality, and can watch the virtual environment from a full angle. If it is necessary to make the virtual objects have intelligence in the virtual space, it is necessary to program them accordingly.

2.4.2. Speech recognition models for cross-cultural communication. (1) *Acoustic modeling.* Acoustic models are one of the core parts of speech recognition algorithms, which are used to convert input speech signals into corresponding acoustic features and predict the corresponding phonemes or words based on these features. Common acoustic models are traditional Gaussian mixture models and deep learning models.

For single Gaussian models, the Gaussian distribution obeys the probability density function below:

$$P(x|\theta) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{the sample is one} - \text{dimensional data}, \\ \frac{1}{(2\pi)^{\frac{\omega}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right), \text{the sample is multidimensional data}, \end{cases} \tag{1}$$

where μ is the data mean, σ is the data standard deviation, Σ is the covariance, and ω is the data dimension.

A Gaussian mixture model can be viewed as a combination of K single Gaussian distributions. Each distribution is a component of the model, and these components act as hidden variables that affect the mixed model as a whole:

$$P(x|\theta) = \sum_{k=1}^{K} \alpha_k \rho(x|\theta_k), \tag{2}$$

$$\theta = (\tilde{\mu}_k, \tilde{\sigma}_k, \tilde{\alpha}_k), \tag{3}$$

where x_j denotes the jth observation, $j = 1, 2, 3..., N$, K is the number of sub-Gaussian models in the mixture model, $k = 1, 2, ..., K$, α_k is the probability that the observation belongs to the kth sub-model and $\rho(x|\theta_k)$ is the Gaussian distribution density function of the kth sub-model.

Its maximum likelihood estimation function is:

$$\log L(\theta) = \sum_{j=1}^{N} \log P(x_j|\theta) = \sum_{j=1}^{N} \log \left(\sum_{k=1}^{K} \alpha_k \rho(x|\theta_k) \right). \tag{4}$$

At this point it is not possible to obtain the maximum parameters of the maximum likelihood estimation function by derivation, which is generally solved iteratively by the EM algorithm. The parameters are first initialized and based on the current parameters, the likelihood of each data j coming from sub-model k is calculated:

$$\varphi_{jk} = \frac{\alpha_k \rho(x_j|\theta_k)}{\sum\limits_{k=1}^{K} \alpha_k \rho(x_j|\theta_k)}. \tag{5}$$

Then, the model parameters for the new iteration are calculated, where φ_{jk} denotes the probability that the jnd observation belongs to the krd sub-model, as shown in the following equation:

$$\mu_k = \frac{\sum\limits_{j=1}^{N} (\varphi_{jk} x_j)}{\sum\limits_{j=1}^{N} \varphi_{jk}}, \tag{6}$$

$$\Sigma_k = \frac{\sum\limits_{j}^{N} \varphi_{jk} (x_j - \mu_k)(x_j - \mu_k)^T}{\sum\limits_{j}^{N} \varphi_{jk}}, \tag{7}$$

$$\alpha_k = \frac{\sum\limits_{j=1}^{N} \varphi_{jk}}{N}. \tag{8}$$

Repeat calculation φ_{jk} with each model parameter until convergence. Thanks to the powerful modeling capability of neural networks, the acoustic model of speech recognition system is gradually shifting from the traditional Gaussian mixture model to the neural network represented by DNN.

A deep neural network is used to replace the GMM acoustic model in traditional GMM-HMM systems, eliminating the need for assumptions on the distribution of acoustic features. Thanks to the DNN architecture, it is possible to process successively spliced frames, which helps to utilize contextual information more efficiently. Before model training begins, each frame of speech is forced to align with the output of the DNN by using Viterbi on the training data. During training, DNN output and input features are usually used, which in turn replaces the GMM with the trained DNN in order to compute observation probabilities. Other components of the model such as transfer probabilities and initial probabilities remain unchanged.The DNN-HMM network structure is shown in Figure 1.

(2) *Pronunciation dictionary.* Pronunciation dictionary, in Chinese, is the correspondence between pinyin and Chinese characters, and in English, is the correspondence between phonetic symbols and words. Generally speaking, the larger the size of the pronunciation dictionary and the more words it

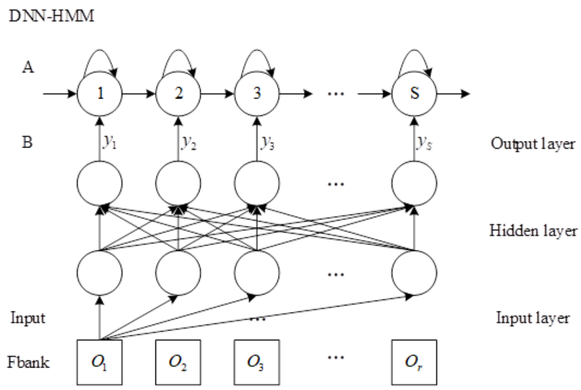

Fig. 1. DNN-HMM network structure

covers, the higher the accuracy of speech recognition, as in life, dictionaries with a small vocabulary are often less accessible than dictionaries with a large vocabulary.

(3) *Language modeling.* The language model can eliminate some combinations that do not meet the language expression habits, thus reducing a large number of paths to be searched and greatly reducing the decoding time. The language model represented by n-gram is widely used in the field of speech recognition due to its simple structure.

The probability of occurrence for a given target sequence $S = \{t_1, t_1, ...t_n\}$ is calculated as shown in Eq. The n-gram assumes that the occurrence of the Nth word is only related to the previous N-1 words:

$$P(S) = P(t_1, t_2, ..., t_n) = P(t_1)P(t_2|t_1)...P(t_n|t_1t_2...t_{n-1}). \tag{9}$$

Increasing the number of parameters corresponding to the conditions in the conditional probability will show an exponential growth, at which point the model will become more complex, so consider the extreme case of retaining only $P(t_i)$, whereupon the above equation will be simplified to:

$$P(S) = P(t_1, t_2, ..., t_n) = \prod_{i=1}^{n+1} P(t_i|t_0t_1...t_{i-1}) \approx \prod_{i=1}^{n} P(t_i). \tag{10}$$

The value of n is judged according to the specific situation, and a common one is the binary Bi-Gram, which means that the probability distribution of the current word is only related to the nearest word, and the formula is shown in (11):

$$P(S) = P(t_1, t_2, ..., t_n) = \prod_{i=1}^{n+1} P(t_i|t_0t_1...t_{i-1}) \approx \prod_{i=1}^{n} P(t_i|t_{i-1}). \tag{11}$$

The ternary Tri-Gram, i.e., the probability distribution of the current word is related to the two words that are closest to each other, is shown in the formula in (12):

$$P(S) = P(t_1, t_2, ..., t_n) = \prod_{i=1}^{n+1} P(t_i|t_0t_1...t_{i-1}) \approx \prod_{i=1}^{n} P(t_i|t_{i-1}t_{i-2}). \tag{12}$$

3. Construction of a virtual reality-based ecological model for language learning

3.1. Model presentation

The ecological model of language learning based on virtual reality technology has been reconstructed from technical support, theoretical support, macro level and micro level, and the ecological model of language learning based on virtual reality technology is shown in Figure 2. The purpose of the proposed model is to allow learners to continuously communicate and cooperate in different levels of scenarios, to continuously feedback and evaluate and gradually achieve learning goals, and to truly realize the safe transition from the shallow water to the deep water of language learning.

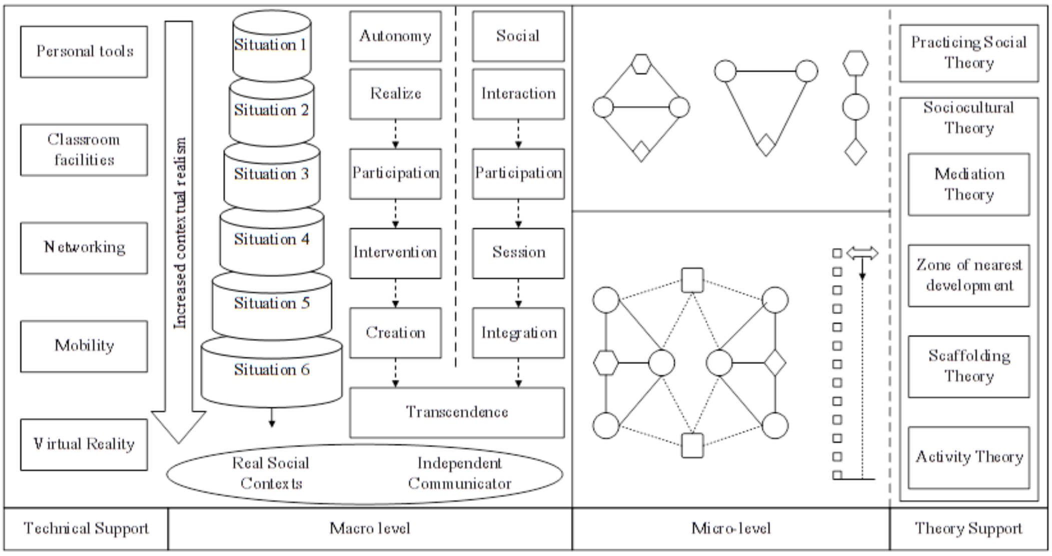

Fig. 2. Language learning ecological model based on virtual reality technology

3.2. Modeling

3.2.1. Technology. The model distinguishes and gives a clear position to virtual reality technology from other forms of technology, which is used to restore real application scenarios and provide an ideal learning place, combining the advantageous features of other technologies, reflecting the advantageous integration of language learning technologies, and being able to utilize unique imaginative and constructive qualities to make the application of environments, roles and strategies in language learning more expressive.

3.2.2. Macro level. Under the background of mass education and the new demand for personalized education, the goal of language learning is more and more inclined to improve independent learning ability and enhance the sense of social interaction. As an individual, independent learning ability, both inside and outside the classroom, should be the key to learners' mastery of a language. However, as a member of a learning community, the quality of language learning is directly related to the effectiveness of learners' social interaction.

We take the context as the unit of language learning activities, and the context is a "microcosm" of the reality of language learning and use, and the result of the virtualization of the relevant real-life situation by technology. Each learning unit here is a micro-representation of the real social context, which compresses the complexity and realism of the actual interaction context. However, with the deepening of language learning, the authenticity of the context in the learning unit will increase, and the social indicators such as the types of social roles, the complexity of social relations, the richness of social activities, and the degree of environmental authenticity in the context will be increased, and eventually converge with the real social context. This form of learning ensures that learners have a real sense of social presence at all times, and the establishment of a hierarchical language learning context is conducive to promoting the smooth transfer of learning.

3.2.3. Micro-level. The micro-level is the concrete realization of the macro-level requirements, reflecting the activities carried out in the context unit. The roles in the model can be divided in two ways. According to the participants' roles, they can be divided into: teachers, students, virtual roles and native speakers, in which the virtual roles are the simulation of teachers', students' and native speakers' behaviors, and the significance of the native speaker's participation is to establish the practice conditions for using the target language and to carry out tandem learning. From the point of view of the role of the activity it can be subdivided into two categories: instructors and learning companions. In addition, although the activities are carried out in a learner-centered way, the model does not discard the role of the teacher as a supervisor and guide for the learners, and emphasizes the guidance and accompaniment between the learners.

4. Cross-cultural contextualized teaching/learning VR system for language learning

Based on the analysis and research of related literature and the theoretical basis and viewpoints of this paper, this paper proposes a VR system applicable to teaching/learning in cross-cultural contexts of language learning as shown in Figure 3. The design concept of this system consists of five types of participants (R&D team, teaching experts, teachers, Chinese learners and native speakers), eight modules (fundamental goal, theoretical foundation, construction principle, technical conditions, construction method, context type, teaching process, learning mode), one engine (open VR development engine) and one system (contextual resource retrieval system) that are jointly organized into a virtual learning space. The fundamental goal of this VR system is to help language learners acquire better language knowledge and improve their language intercultural communication skills while having fun.

5. Case studies

5.1. Speech interaction recognition performance

In order to verify the recognition performance of the speech recognition model for cross-cultural communication proposed above, a set of experiments is done below for verification.

In this experiment, 15 students are selected as testers (numbered 1∼15) to communicate cross-culturally (spoken) in the language learning system constructed in this paper. The voice interaction time thresholds were all set to 2500 ms. the testers in this experiment used voice interaction to complete the input. Effective input is the number of words that were effectively recorded into the device after the testers completed the voice interaction, and the effective input rate is the data obtained by dividing the number of effective input words by the total number of words in the

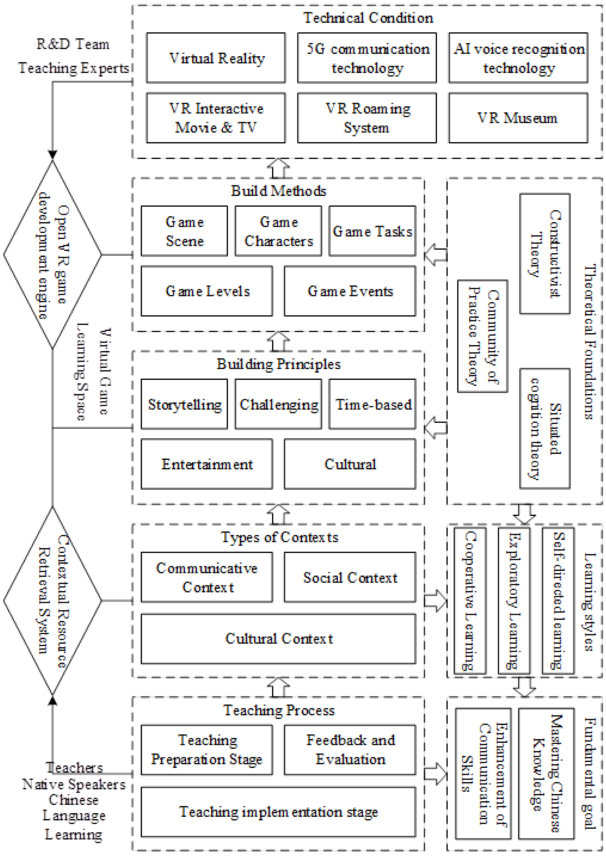

Fig. 3. Cross-cultural situational teaching VR system in language learning

experiment multiplied by one hundred percent. Correct input is the number of words correctly operated in the effective input, and the correct input rate is the data obtained by dividing the number of correctly input words by the number of validly input words multiplied by one hundred percent. The accuracy of the speech recognition model for cross-cultural communication is verified by the effective input rate and correct input rate. The experimental results and data of this experiment are shown in Table 1.

Observing the experimental data in Table 1, it is found that the effective input rate and correct input rate of the cross-cultural communication speech recognition model in this paper are 99.7% and 99.5%, respectively, which are both more than 99%, which indicates that the cross-cultural communication speech recognition model in this paper has excellent recognition rate and correct rate. In addition, the effective input rate and correct input rate of more than 99% can basically meet the requirements of cross-cultural communication in virtual simulation situations.

5.2. Language learning interactions

In order to test the impact of the intercultural contextualized teaching/learning VR system on students' language learning interactions in this paper. In this section, the data collected from students in the intercultural contextual teaching/learning VR system in three intercultural communication lessons will be analyzed for the ratio of classroom interaction behaviors in order to find out the facilitating effect of the intercultural contextual teaching/learning VR system on classroom interaction. In the interactive system, students are language learners and teachers are native speakers of a foreign language.

Table 1. Speech interaction experiment results

Number	Input word	Effective input word	Effective input rate	Correct input word	Correct input rate
1	700	700	100%	700	100%
2	739	739	100%	732	99%
3	716	716	100%	716	100%
4	638	638	100%	625	98%
5	547	542	99%	542	100%
6	539	528	98%	517	98%
7	762	762	100%	762	100%
8	714	714	100%	714	100%
9	532	532	100%	532	100%
10	792	792	100%	784	99%
11	523	523	100%	523	100%
12	539	534	99%	534	100%
13	612	606	99%	606	100%
14	764	764	100%	764	100%
15	519	519	100%	514	99%
Mean	642.4	640.6	99.7	637.7	99.5

5.2.1. Analysis of classroom structure. In order to better analyze the classroom structure as a whole, the author extracted the ratios of teacher's speech, student's speech, teacher's indirect influence, and teacher's direct influence from the ratio table of intercultural communication classroom interaction behaviors in the three cases, and the classroom structure ratios are shown in Table 2.

In a regular classroom, teacher verbal activity is much more active than student verbal activity, with a teacher verbal norm of 68% and a student verbal norm of 20%, a ratio of 3.4:1. Observation of Table 2 reveals that there is relatively less teacher verbal activity and relatively more student verbal activity in all three lesson examples. The ratios of teacher speech in the three intercultural communication classrooms were 42.48%, 40.09%, and 41.69%, respectively, and the ratios of student speech were 40.89%, 41.99%, and 42.46%, respectively, with student speech and teacher speech being comparable. Compared with the traditional classroom, the intercultural communication classroom scenario in the cross-cultural contextual teaching/learning VR system allows students to express themselves more, and the teacher actively guides students to respond, participate, and discuss cooperatively, making the classroom interaction more active.

Table 2. Class structure ratio

Code	Class 1	Class 2	Class 3
Teacher speech	42.48%	40.09%	41.69%
Student speech	40.89%	41.99%	42.46%
Teacher indirect influence	23.44%	25.06%	29.54%
Teacher direct influence	19.04%	15.03%	12.15%

Teachers' direct and indirect influences reflect the way teachers control students in classroom teaching. When the rate of indirect influence is higher than that of direct influence, it means that teachers tend to use indirect influence to guide students, and the classroom teacher-student relationship is harmonious. In all three cases, teachers' indirect influence was higher than teachers' direct influence, and teachers tended to use indirect influence to control students. Teachers' direct influence rate decreased in the last two cases, and teachers consciously reduced their unilateral control over students, and made students feel involved in the classroom by encouraging them to agree and asking questions.

5.2.2. Teacher impact analysis. The indirect influence part of the teacher is composed of X1 "positive response from the teacher", X2 "adopting students' opinions", X3 "asking closed questions", X4 "asking open questions", and the direct influence part of the teacher is composed of X5 "teaching", X6 "instruction", X7 "guiding and inspiring", and X8 "negative response from the teacher".

Table 3 shows that in the three intercultural communication classes, the indirect influence of teachers increased from 23.44% to 29.54%, and the direct influence of teachers decreased from 19.04% to 12.15%. The proportion of indirect influence behaviors of teachers such as X1 "positive response of teachers", "positive response of students" of X2, "asking closed-ended questions" of X3 and "asking open-ended questions" of X4 all showed an upward trend, and in the direct influence of teachers, except for X7 "guidance and inspiration", the rates of X5 "teaching", X6 "instruction" and X8 "negative response of teachers" all showed a downward trend. In the cross-cultural communication of the VR system for cross-cultural teaching/learning, the direct influence of the teacher, who is a native speaker of a foreign language, gradually weakens, encourages students' active expression, and enriches the learning interaction in cross-cultural communication.

Table 3. Teacher indirect and direct influence ratio

	Total ratio		
	Class 1	Class 2	Class 3
Teacher indirect influence	23.44%	25.06%	29.54%
X1	6.48%	7.05%	8.95%
X2	7.42%	7.89%	9.15%
X3	2.33%	2.86%	3.12%
X4	7.21%	7.26%	8.32%
Teacher direct influence	19.04%	15.03%	12.15%
X5	10.26%	9.12%	7.85%
X6	3.52%	3.09%	1.23%
X7	3.65%	1.77%	2.39%
X8	1.61%	1.05%	0.68%

5.3. Analysis of language learning effectiveness

A semester-long teaching experiment was conducted with 100 randomly selected students from School Y, divided into an experimental group and a control group, in which the experimental group used the Cross-Cultural Contextual Teaching/Learning VR System while the control group followed the traditional mode of learning English. This section presents the results of the pre-test and post-test of the experimental and control groups. Before the experiment, both groups took a pre-test to find out the actual English level of the students and to verify the similarity between the experimental and control teaching methods. After one semester of English learning, a post-test was conducted in both classes to verify the positive effect of the "Cross-Cultural Contextual Teaching/Learning VR System" on students' English proficiency.

5.3.1. Comparison of pre-test results between the two groups. In order to understand the specific effects of the cross-cultural contextualized teaching/learning VR system on students' English proficiency (English thinking ability, English expression ability, cross-cultural communication ability, and cultural comprehension), this study first tested the two groups of students before the experiment to ensure that there was not much difference between the experimental group and the control group in terms of various English proficiencies in order to compare the two groups at a later stage and highlight the advantages of the cross-cultural contextualized teaching/learning VR system. The English proficiency statistics of the students in the experimental group and the control group were tested using SPSS 27.0 software for paired samples. The test results are shown in Table 4.

As shown in Table 4, this study used independent samples t-test to analyze the comparison between the experimental and control groups in the pre-test, and found that the comparison coefficients between the experimental and control groups in the pre-test for each English proficiency are as follows.

Table 4. Comparative analysis of pre-test result independent sample t-test

Variable	Group	N	Mean	SD	t	P
English thinking ability	Experimental group	50	4.526	0.412	0.745	0.496
	Control group	50	4.489	0.405		
English expressing ability	Experimental group	50	5.985	0.623	-0.142	0.965
	Control group	50	5.997	0.618		
Cross-cultural communication ability	Experimental group	50	3.568	0.522	0.563	0.674
	Control group	50	3.482	0.589		
Culture understanding ability	Experimental group	50	4.673	0.842	0.476	0.712
	Control group	50	4.584	0.853		
Total	Experimental group	50	18.752	2.052	0.549	0.722
	Control group	50	18.552	1.976		

The t-values of English thinking ability, English expression ability, intercultural communication ability, intercultural communication ability and total score are 0.745, -0.142, 0.563, 0.476, 0.549, respectively, and the corresponding p-values are 0.496, 0.965, 0.674, 0.712, 0.722, respectively, which are all greater than 0.05.In summary, according to the results of independent samples t-test, there is no significant difference between the experimental and control groups' pre-tests in terms of English thinking ability, English expression ability, intercultural communication ability, intercultural communication ability and total score.

5.3.2. Comparison of posttest results between the two groups. In order to verify whether the cross-cultural contextual teaching/learning VR system has changes on students' English proficiency, after one semester, the author conducted a post-test on the experimental and control groups, and the specific analysis results are shown in Table 5.

There is a significant difference between the experimental and control groups' post-test scores in English thinking ability, English expression ability, intercultural communication ability, intercultural communication ability and total score. The mean values of the experimental group in the five areas were 8.845, 9.662, 9.074, 8.975, and 36.556, and the mean values of the control group were 4.875, 5.692, 3.482, 5.015, and 19.064, respectively. t-values of the independent samples t-tests showed that the t-values were 5.694, 6.894, 4.628, and 2.937, respectively, 7.488, with p-values less than 0.05,

Table 5. Comparative analysis of post-test result independent sample t-test

Variable	Group	N	Mean	SD	t	P
English thinking ability	Experimental group	50	8.845	0.395	5.694	0.005
	Control group	50	4.875	0.462		
English expressing ability	Experimental group	50	9.662	0.674	6.894	0.004
	Control group	50	5.692	0.634		
Cross-cultural communication ability	Experimental group	50	9.074	0.411	4.628	0.001
	Control group	50	3.482	0.569		
Culture understanding ability	Experimental group	50	8.975	0.823	2.937	0.007
	Control group	50	5.015	0.945		
Total	Experimental group	50	36.556	3.585	7.488	0.000
	Control group	50	19.064	2.641		

indicating that there are significant differences between the two groups, the experimental group and the control group, in the four dimensions of English proficiency and the total score.

In summary, according to the results of independent samples t-test, there are significant differences between the experimental and control groups in all variables (English thinking ability, English expression ability, intercultural communication ability, intercultural communication ability and total score). The experimental group's posttest scores in these areas were significantly better than those of the control group.

5.3.3. Comparison of pre- and post-tests in the experimental group. In order to understand the specific effects of the VR system of teaching/learning in cross-cultural contexts on students' four English language proficiencies, this study collected statistical data on the statistical changes in the performance of the experimental group of students in each dimension. The results of the study were tested using SPSS 27.0 software for paired samples. The results of the analysis are shown in Table 6.

As shown in Table 6, the results of the comparison of each English proficiency in the pre and post-tests conducted for the experimental group show the following: the post-test means of the experimental group in English thinking ability, English expression ability, intercultural communication ability, intercultural communication ability and the total scores were 8.845, 9.662, 9.074, 8.975 and 36.556 respectively, and the pre-test means were 4.526, 5.985, 3.568, 4.673, 18.752. The paired differences were -4.319, -3.677, -5.506, -4.302, -17.804. The paired samples t-test found that the p-values were all less than 0.05, which indicated that there was a significant difference between the pre and post-test scores of the experimental group's English language ability.

According to the results of the paired samples t-test, there is a significant difference between the pre and post-tests of English proficiency in the experimental group, specifically the post-test scores of the experimental group are significantly better than the pre-test scores in terms of English Thinking Ability, English Expression Ability, Intercultural Communication Ability, Intercultural Communication Ability and the Total Score. This indicates that the experimental group has significant improvement in English proficiency, and the cross-cultural contextual teaching/learning VR system can help users with good language learning results.

Table 6. Comparison of pre-test and post-test English abilities of experimental group

Variable	Pre/post	N	Mean	SD	Pair difference	t	P
English thinking ability	Pre-test	50	4.526	0.412	-4.319	-9.484	0.002
	Post-test	50	8.845	0.395			
English expressing ability	Pre-test	50	5.985	0.623	-3.677	-10.265	0.003
	Post-test	50	9.662	0.674			
Cross-cultural communication ability	Pre-test	50	3.568	0.522	-5.506	-8.946	0.001
	Post-test	50	9.074	0.411			
Culture understanding ability	Pre-test	50	4.673	0.842	-4.302	-7.458	0.002
	Post-test	50	8.975	0.823			
Total	Pre-test	50	18.752	2.052	-17.804	-15.165	0.000
	Post-test	50	36.556	3.585			

6. Conclusion

On the basis of contextual teaching method, the author uses virtual reality technology to create cross-cultural communication situations and speech recognition models in language learning, constructs a virtual reality-based ecological model of language learning, and designs a cross-cultural contextual teaching/learning VR system for language learning according to the language learning model. After an example study, the speech interaction recognition performance, learning interaction situation and application effect of the language learning system in this paper are tested.

The cross-cultural context teaching/learning VR system has an effective language recognition rate of 99.7%, a correct rate of 99.5%, and a high speech recognition accuracy.

In the cross-cultural communication classroom scenarios of the language learning system in this paper, the teacher speech rate is 42.48%, 40.09%, 41.69%, and the student speech rate is 40.89%, 41.99%, 42.46%, which are comparable. Teachers' indirect influence increased from 23.44% to 29.54% and direct influence decreased from 19.04% to 12.15%. Learning interactions in the intercultural communication classroom were more positive.

The p-values of the pre-test English thinking ability, English expression ability, intercultural communication ability, intercultural communication ability and total scores of the experimental group and the control group are all greater than 0.05. The pre-test English ability of the two groups is not different. The mean values of the post-test English proficiency of the experimental group were 8.845, 9.662, 9.074, 8.975, and 36.556, which were higher than those of the control group by 3.970, 3.970, 5.592, 3.960, and 17.492, respectively. p-values were all less than 0.05, and the two groups' post-test scores showed significant differences. The difference between the pre and post-test scores of the experimental group is 4.319, 3.677, 5.506, 4.302, 17.804 respectively, the post-test is much higher than the pre-test, and the p-values are all less than 0.05, and the experimental group's English language proficiency has been significantly improved.

About the Author

Li Chen, 1980.2, born in Tongxu County, Henan Province, a Bachelor of Arts degree from Xinyang Normal University in June 2003, a Master of Arts degree from Henan University in June 2010, now is an associate professor at Pingdingshan Vocational And Technical College, with his research interests: linguistics and applied linguistics.

References

[1] K. Abrams, D. McMahon, H. Whittenburg, L. Bruno, and J. Firestone. Using virtual reality to teach science vocabulary to transition aged youth with intellectual and developmental disabilities. *Journal of Special Education Technology*, 39(4):527–541, 2024. https://doi.org/10.1177/01626434241239922.

[2] M. H. Ahtif and N. Gandhi. The role of language in cross cultural bonds. *Journal of Asian Multicultural Research for Social Sciences Study*, 3(4):7–16, 2022. https://doi.org/10.47616/jamrsss.v3i4.321.

[3] M. Alfadil. Effectiveness of virtual reality game in foreign language vocabulary acquisition. *Computers & Education*, 153:103893, 2020. https://doi.org/10.1016/j.compedu.2020.103893.

[4] H. Alqudah and M. A. S. Khasawneh. Cross-cultural semantic analysis in online language learning platforms: insights from user-generated content. *Migration Letters*, 21(S3):27–41, 2024.

[5] M. Baralt, S. Doscher, L. Boukerrou, B. Bogosian, W. Elmeligi, Y. Hdouch, J. Istifan, A. Nemouch, T. Khachatryan, and N. Elsakka. Virtual tabadul: creating language-learning community through virtual reality. *Journal of International Students*, 12(S3):168–188, 2022. https://doi.org/10.32674/jis.v12iS3.4638.

[6] P. Bawa. Self-regulation, co-regulation, and feedback in the context of cross-cultural language acquisition in higher education: a conceptual approach. *Journal of Research Initiatives*, 4(1):9, 2018.

[7] A. Berns and S. Reyes-Sánchez. A review of virtual reality-based language learning apps. *RIED Revista Iberoamericana de Educación a Distancia*, 24(1):158–170, 2021. https://doi.org/10.5944/ried.24.1.27486.

[8] A. Cheng, L. Yang, and E. Andersen. Teaching language and culture with a virtual reality game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 541–549, 2017. https://doi.org/10.1145/3025453.3025857.

[9] D. Dobrovol'skij and E. Piirainen. *Figurative Language: Cross-cultural and Cross-Linguistic Perspectives*, volume 350. Walter de Gruyter GmbH & Co KG, 2021.

[10] X. HOU and D. LIU. On the application of situational teaching method in foreign language class from the perspective of interaction. *Higher Education of Social Science*, 1:43–47, 2024. http://dx.doi.org/10.3968/13549.

[11] J. House and D. Kádár. *Cross-Cultural Pragmatics and Foreign Language Learning*. Edinburgh University Press, 2024. https://doi.org/10.1515/9781399523240.

[12] Y.-J. Lan. Immersion into virtual reality for language learning. *Psychology of Learning and Motivation*, 72:1–26, 2020. https://doi.org/10.1016/bs.plm.2020.03.001.

[13] J. Legault, J. Zhao, Y.-A. Chi, W. Chen, A. Klippel, and P. Li. Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages*, 4(1):13, 2019. https://doi.org/10.3390/languages4010013.

[14] Y. Li. The application of situational teaching method in middle school english reading classes. *Education Reform and Development*, 6(7):192–197, 2024. https://doi.org/10.26689/erd.v6i7.7744.

[15] A. Parmaxi. Virtual reality in language learning: a systematic review and implications for research and practice. *Interactive Learning Environments*, 31(1):172–184, 2023. https://doi.org/10.1080/10494820.2020.1765392.

[16] A. Parmaxi and A. A. Demetriou. Augmented reality in language learning: a state-of-the-art review of 2014–2019. *Journal of Computer Assisted Learning*, 36(6):861–875, 2020. https://doi.org/10.1111/jcal.12486.

[17] J. A. Roberto. Cross-cultural languages behind technology-enhanced language learning. In *Cross-Cultural Perspectives on Technology-Enhanced Language Learning*, pages 16–46. IGI Global, 2018. 10.4018/978-1-5225-5463-9.ch002.

[18] T.-Y. Tai, H. H.-J. Chen, and G. Todd. The impact of a virtual reality app on adolescent efl learners' vocabulary learning. *Computer Assisted Language Learning*, 35(4):892–917, 2022. https://doi.org/10.1080/09588221.2020.1752735.