# The Design and Practice of Cross-Cultural Communication Virtual Teaching Scene from the Perspective of Multimodal Fusion

## Chenlu He[1,*], Hui Jia[2], Kaimei Luo[3]

[1]School of International Chinese Education, Beijing Normal University, Beijing, China

[2]Chinese Language Center, Shenzhen MSU-BIT University, Shenzhen, China

[3]Zhongshan Huachen Experimental Middle School, Zhongshan, China

*Corresponding author: hechenluuuuu010616@163.com

**Abstract:** In the context of the deepening development of globalization, international Chinese education has become an important bridge to promote cross-cultural exchanges, and cultivating cross-cultural communication skills is a core goal of international Chinese education. Starting from the perspective of multimodal integration, this paper focuses on the design and practice of virtual teaching scenarios for cross-cultural communication in international Chinese education. First, it examines the relevant theories of multimodal integration, cross-cultural communication, and virtual teaching scenarios, analyzes the current situation and challenges of cross-cultural communication in international Chinese education, and discusses the application advantages of multimodal integration in this context. Based on this, this paper proposes design principles and elements for virtual teaching scenarios that facilitate cross-cultural communication, and analyzes them in conjunction with specific cases. Finally, the conclusions are summarized, and the limitations and future directions of the research are highlighted, aiming to provide a reference for enhancing the quality of cross-cultural communication teaching in international Chinese education.

## 1. Introduction

In the context of globalization, international Chinese education, as a vital carrier of cultural dissemination and cross-cultural communication, has become increasingly significant. The rapid rise of multimodal fusion technology in the field of education has opened up new possibilities for optimizing teaching models. This study aims to design cross-cultural communication virtual teaching scenarios from the perspective of multimodal fusion and explore their application in international Chinese education. This study not only enriches the theoretical framework of international Chinese education but also provides new methodological guidance for cross-cultural communication teaching, thereby improving the language application ability and cultural adaptation ability of Chinese learners at the practical level. The study will comprehensively employ the literature research method to identify relevant theories, utilize the case analysis method to examine the advantages and disadvantages of existing practice models, and employ the survey research method to collect feedback data from learners and teachers. Its innovation is mainly reflected in the in-depth combination of multimodal fusion theory and virtual teaching scenarios, which constructs a cross-cultural communication teaching framework suitable for international Chinese education, and proposes an operational scenario design strategy.

## 2. Theoretical Basis

### 2.1 The Core Connotation and Educational Application of Multimodal Fusion Theory

Multimodal fusion refers to the process of transmitting information through the synergy of multiple sensory modalities such as vision, hearing, touch, and kinesthetics. In this process, various

modalities cooperate and complement each other to jointly construct a comprehensive system.

In the field of education, multimodal fusion has important application value. It can present knowledge in a multi-dimensional manner and cater to the diverse needs of different learners. For example, for visual learners, visual modalities such as images and videos can help them understand knowledge more quickly. For auditory learners, auditory modalities such as audio explanations and dialogues are more effective. Different modal combinations have different adaptability in teaching. For example, the combination of "text + image + audio" is often used for systematic explanation of knowledge. Text provides precise conceptual expression, images assist in understanding abstract content, and audio enhances the appeal of information; the combination of "virtual scene + interactive operation" is more suitable for practical teaching, allowing learners to master skills through personal experience [1].

## 2.2 Theoretical Support for Cross-Cultural Communication and Virtual Teaching

There are many core theories of cross-cultural communication. Hofstede's cultural dimension theory analyzes the differences between cultures along the dimensions of power distance, individualism and collectivism, masculinity and femininity, uncertainty avoidance, long-term orientation, and short-term orientation, providing an important perspective for understanding behavioral differences in cross-cultural communication. The high- and low-context culture theory categorizes culture into two distinct types: high-context culture and low-context culture. The transmission of information in high-context cultures depends more on the context, while in low-context cultures, it depends more on clear language expression. Cross-cultural communication ability encompasses cultural cognition, effective communication strategies, and empathy. Cultural cognition is the foundation, enabling learners to understand the characteristics and differences of various cultures. Communication strategies serve as a guarantee, helping learners navigate different situations in actual communication. Empathy facilitates a deeper understanding of the other party's emotions and intentions [2].

The theoretical basis for constructing virtual teaching scenarios includes situational learning theory and experiential learning theory. The situational learning theory emphasizes that learning should be carried out in real situations, and learners acquire knowledge and skills through interaction with the environment and participation in social practice; the experiential learning theory believes that learning is a cyclical process through experience, reflection, abstraction and application, allowing learners to deepen their understanding of knowledge through personal experience. Virtual reality (VR), augmented reality (AR), mixed reality (MR), and other technologies have unique application logic in scene creation. VR can create a completely virtual immersive environment for learners, making them feel as if they are in a real scene; AR superimposes virtual information on the real environment to achieve a combination of virtual and real; MR further integrates the advantages of VR and AR, allowing virtual objects to interact more naturally with the real environment.

## 3. The Current Situation and Challenges of Cross-Cultural Communication in International Chinese Education

### 3.1 The Development Trend of International Chinese Education and Cross-Cultural Teaching

International Chinese education has evolved from instrumental teaching to culturally inclusive teaching. Early international Chinese education focused on imparting knowledge of the Chinese language, to develop learners' Chinese listening, speaking, reading, and writing skills, and emphasized the practicality of Chinese as a communication tool. With the development of globalization, people have increasingly come to recognize the significance of culture in cross-cultural communication. International Chinese education has begun to shift towards culturally inclusive teaching, which not only teaches language knowledge but also focuses on cultivating learners' understanding and respect for Chinese culture, as well as cross-cultural communication skills [3].

Currently, the scope of international Chinese education is expanding, and the learner group is becoming increasingly diverse, encompassing individuals of various ages, occupations, and cultural

backgrounds. The cultivation of cross-cultural communication skills is at the core of the goal system of international Chinese education. It is closely related to the cultivation of language ability and cultural literacy and develops in a coordinated manner. Language ability is the foundation of cross-cultural communication. Without a solid language foundation, it is difficult to communicate effectively; cultural literacy is the support of cross-cultural communication, allowing learners to show their understanding and respect for different cultures in communication; and cross-cultural communication ability is a comprehensive reflection of language ability and cultural literacy, and is one of the ultimate goals of international Chinese education.

## 3.2 Dilemma of Cross-Cultural Communication Teaching

In terms of teaching methods, teachers often spend a considerable amount of time explaining cultural knowledge and communication rules, but provide learners with few opportunities to engage in actual communication practices. At the same time, culture is fragmented, and only some cultural phenomena are introduced sporadically in class, lacking systematicity and depth, which makes it difficult for learners to form a holistic understanding of culture.

In terms of technology application, the integration of multimodal resources is insufficient, and various teaching resources such as text, images, audio, and video are often used in isolation, without forming a synergistic effect. The interactivity and authenticity of virtual scenes need to be improved. Some virtual scenes are simply screen displays, and learners cannot interact effectively with them. In addition, the cultural elements and communication situations in the scenes are far removed from real-life situations, which affects the teaching effect [4].

In terms of effect evaluation, there is a lack of a dynamic, multi-dimensional evaluation system for cross-cultural communication competence. Current evaluations often focus on learners' memory of cultural knowledge and language test scores. Still, there is a lack of evaluation of learners' performance in actual cross-cultural communication, such as the use of communication strategies and cultural adaptability, which makes it difficult to fully reflect learners' cross-cultural communication competence.

## 4. Advantages of Multimodal Fusion in International Chinese Education

### 4.1 Optimize the Way of Chinese Knowledge Transfer and Cultural Perception

Multimodal fusion can make abstract language knowledge concrete and complex cultural connotations layered. For example, in pronunciation teaching, videos are used to demonstrate the movement trajectory of the pronunciation organs, so that learners can intuitively understand pronunciation skills; in grammar teaching, animations are used to analyze the application scenarios of grammar rules, making abstract grammatical concepts vivid and easy to understand; in cultural teaching, virtual scenes are used to restore cultural customs, allowing learners to experience the charm of Chinese culture in an immersive way.

This multimodal knowledge transfer method will help learners establish the connection between language symbols and cultural meanings. When learners see the word "dumplings", they can not only understand its writing and pronunciation through text, but also see the appearance of dumplings through images, watch the process of making dumplings through videos, and experience eating dumplings through virtual scenes, to deeply analyze the cultural significance of dumplings and improve the efficiency of knowledge reception [5].

### 4.2 Stimulating Immersion in Cross-Cultural Communication Practice

Multimodal virtual scenarios can create communication situations that closely resemble reality, such as simulated business negotiations, daily greetings, and festival celebrations. In these scenarios, learners can participate in the interaction through role-playing and task completion. For example, in a business negotiation scenario, learners assume different roles: one represents a Chinese company, and the other represents a foreign company. They need to use the Chinese knowledge they have learned and cross-cultural communication strategies to negotiate.

In this interactive process, learners can personally engage with communication methods and etiquette across various cultural backgrounds, mastering effective communication skills and resolving potential cultural conflicts. This type of practical learning can enhance learners' motivation and emotional investment, enabling them to transition from passively accepting knowledge to actively participating in the learning process, and thereby improve their enthusiasm and initiative.

## 5. Principles and Elements of Cross-Cultural Communication Virtual Teaching Scene Design

### 5.1 Core Principles of Scene Design

#### 5.1.1 Modal Synergy Principle

In the design of virtual teaching scenes, it is essential to ensure that visual, auditory, interactive, and other modalities work together to avoid information redundancy or loss, thereby forming a unified system for meaning construction. In terms of vision, the scene pictures should be clear and realistic, accurately reflecting the characteristics and cultural elements of the scene, and the text logos should be concise and clear, matching the content of the picture; in terms of hearing, the voice should be standard and natural, and the background sound effects should be consistent with the atmosphere of the scene to enhance the realism of the scene; in terms of interaction, the operation feedback should be timely and accurate, and the role should be consistent with its identity and situation, so that learners can interact with the virtual scene smoothly.

#### 5.1.2 Principles of Cultural Authenticity and Communicative Practicality

The cultural elements in the scene, such as clothing, etiquette, values, etc., need to be designed based on real cultural prototypes to avoid stereotypes. For example, when designing a traditional Chinese wedding scene, it is necessary to accurately represent traditional wedding costumes and ceremony processes, so that learners can understand authentic Chinese traditional culture. In addition, the design of communication tasks should align with the actual needs of learners, such as studying abroad, working, or traveling, to enhance the practical value of the scene. If learners are primarily learning Chinese for studying abroad, then relevant scenes, such as campus life and classroom exchanges, can be designed.

### 5.2 Key Elements of Scene Composition

#### 5.2.1 Virtual Environment and Roles

It is necessary to build a diversified scenario library covering various aspects of life, education, and the workplace, such as "Chinese family dinner" and "cross-border business meeting." The "Chinese family dinner" scenario can help learners understand Chinese family culture and table manners; the relevant scenarios of "cross-border business meeting" can help cultivate learners' cross-cultural communication skills in business occasions.

The role setting must reflect cultural diversity, including virtual communication objects of different ages, occupations, and cultural backgrounds, and their language style and behavior patterns must conform to the cultural characteristics of these groups. For instance, older roles might prioritize traditional etiquette and use more formal language, whereas younger roles tend to be more vibrant and employ colloquial expressions.

#### 5.2.2 Design of Multimodal Task Chain

Based on language and cultural goals, a step-by-step task system is constructed, presenting the characteristics of progressive learning from shallow to deep. The primary task focuses on applying and consolidating language points. Using the restaurant scene as an example, designing the ordering task helps learners skillfully use Chinese quantifiers, such as "a bun" and "a bowl of noodles," to strengthen their basic language knowledge.

The intermediate tasks focus on deepening cultural cognition. In festive scenes, by analyzing the semantic connotations of festive greetings, such as the cultural implications of "Congratulations on your prosperity" in the context of the Spring Festival, learners can establish connections between

cultural symbols and meanings, thereby enhancing their understanding of cultural phenomena.

Advanced tasks focus on resolving cross-cultural conflicts. In business negotiations and other scenarios, communication disagreements caused by cultural differences are designed to guide learners in using appropriate communication strategies to resolve conflicts and develop their problem-solving skills in complex cross-cultural environments.

During the task implementation process, a multimodal support mechanism is embedded. Text prompts help learners reinforce their memory of relevant language knowledge and task requirements. Voice guidance offers real-time error correction and standardization of learners' pronunciation and expression. Animation demonstrations present task operation processes and key precautions intuitively, forming a multidimensional learning support system.

## 6. Case Study on the Practice of Cross-Cultural Communication Virtual Teaching Scenario from the Perspective of Multimodal Integration

### 6.1 VR Immersive Scene: The Experience Course of China Traditional Solar Terms Culture

The "Spring Outing during Qingming Festival" and "Making Dumplings on the Winter Solstice" scenes, built with VR technology, have created a highly immersive Chinese traditional solar term cultural experience environment. In the "Spring Outing during Qingming Festival" scene, learners enter a virtual outing scene in the suburbs through VR equipment. The scene restores the natural landscape of blooming flowers and lush vegetation through high-definition visual presentation technology. At the same time, it superimposes a three-dimensional sound system to simulate the sounds of birds singing and people talking, creating a realistic scene atmosphere close to reality. The scene features a built-in virtual tour guide character, whose function is to systematically explain the folk origins, historical evolution, and related cultural symbols of the spring outing during Qingming Festival. Learners can interact with the virtual tour guide through the interactive interface. In this process, learners have acquired solar term-related vocabulary, such as "tomb sweeping", "outing", and "kite", and deepened their understanding of the cultural connotation of the Qingming Festival, which emphasizes "remembering ancestors and getting close to nature", through scene experience [6].

The "Making Dumplings on the Winter Solstice" scene uses motion capture technology to achieve real-time interaction between learners and the virtual operation interface. Learners can complete operations such as filling preparation and cooking in a virtual environment. The virtual guidance character simultaneously explains the folk origins and operation specifications of eating dumplings on the Winter Solstice based on the preset cultural script. Following the implementation of the course, a combination of quantitative and qualitative evaluation methods revealed that learners' scores on the cultural identity scales were significantly improved, and the cognitive dimension of Chinese traditional solar term culture was expanded. At the language application level, the accuracy of vocabulary use related to the solar terms and adaptability to the context were significantly improved [7].

### 6.2 AR Augmented Reality: Training in Chinese Daily Communication Situations

The application of AR in scenarios such as "supermarket shopping" and "hospital consultation" has built a virtual-real Chinese daily communication training system. In the "supermarket shopping" scenario, learners scan the physical supermarket shelf cards through mobile terminals to trigger the augmented reality presentation of the virtual supermarket scene and the salesperson role. The system generates dialogue tasks based on preset communication scripts, guiding learners to complete communication behaviors such as asking about product prices (e.g., "How much is this apple per pound?") and expressing purchase intentions (e.g., "I want to buy two pounds of apples").

The "Hospital Consultation" scenario activates the virtual doctor role by scanning the physical hospital scene card. Learners need to use Chinese to complete communication tasks, such as describing symptoms (e.g., "I have a headache" and "I have a fever") based on the virtual symptoms they have set. The virtual doctor role will conduct consultations and responses based on the preset diagnosis and treatment logic. Practical data show that this training model significantly improves

learners' Chinese oral fluency, and the naturalness and accuracy of their language expression are improved. Additionally, at the level of communication strategies, learners have significantly improved their mastery of communication skills in specific situations, such as shopping, bargaining, and describing symptoms.

## 7. Conclusion and Outlook

### 7.1 Conclusion

The design logic of the cross-cultural communication virtual teaching scene from the perspective of multimodal fusion is clear, that is, the scene positioning is determined based on the relevant theoretical support of multimodal fusion, cross-cultural communication and virtual teaching, the design goals are clarified in combination with the current situation of international Chinese education. The scene elements are constructed according to the principles of modal collaboration, cultural authenticity, and communicative practicality.

This scenario has significant practical value in international Chinese education. It will effectively enhance learners' interest in cross-cultural communication, allowing them to participate more actively in cross-cultural communication learning. Moreover, it enhances learners' understanding of Chinese culture, allowing them to not only master Chinese language knowledge but also truly understand the cultural connotations behind the language.

### 7.2 Research Limitations and Future Trends

This study has certain limitations. The technical stability of some virtual scenes needs to be optimized, as problems such as interaction delays occasionally occur during use, which negatively impact learners' experience.

Future research can be carried out from the following aspects: First, it is suggested to explore the combination of artificial intelligence (AI) and multimodal fusion technology to achieve intelligent responses of virtual characters and dynamic adjustment of scenes, make virtual scenes more intelligent and adaptable, and be able to adjust teaching content and difficulty in real time according to learners' learning conditions and needs. Second, it is recommended that a personalized scene recommendation system be developed, tailored to learners from diverse cultural backgrounds. This system would suggest the most suitable virtual teaching scenes based on their cultural background, learning goals, and interests, thereby enhancing overall learning efficiency. Third, it is essential to enhance the long-term evaluation mechanism for cross-cultural communication abilities by integrating learning analytics technology. It will track learners' educational trajectories and the progression of their cross-cultural communication skills, providing a more accurate foundation for teaching improvements.

## References

[1] Shi X, Li X, Toda T. Multimodal Fusion of Music Theory-Inspired and Self-Supervised Representations for Improved Emotion Recognition[C]//Interspeech 2024. 2024. DOI:10.21437/Interspeech.2024-2350.

[2] Okamoto M, Isbister K, Nakanishi H. Supporting Cross-Cultural Communication with a Large-Screen System[J]. New Generation Computing, 2002, 20(2):165-185. DOI:10.1007/BF03037577.

[3] Zhang X .Innovation of teaching methods of international Chinese language education in universities based on artificial intelligence analysis technology[J].Applied Mathematics and Nonlinear Sciences, 2024, 9(1).DOI:10.2478/amns.2023.2.00336.

[4] Fengrong J. Promoting the International Dissemination of Chinese Culture Through International Chinese Language Education: A Case Study of Chinese-English Idiomatic Equivalence[J]. Contemporary Social Sciences, 2023, 8(6):119-140. DOI:10.19873/j.cnki.2096-0212.2023.06.007.

[5] Xiaoying W. Innovation of the Chinese International Education Training Model Based on

Situational Creation[J]. IEEE, 2020.DOI:10.1109/CIPAE51077.2020.00084.

[6] Qingqing K, Lishen Y. The Application of Education for International Understanding of Chinese Language Teaching: A Critical Study[J]. British Journal of Teacher Education & Pedagogy, 2023, 2(1). DOI:10.32996/bjtep.2023.2.1.3.

[7] Meng L , Pan W .Promoting Language and Literature in Colleges and Universities for the Development of International Chinese Language Education Professionals[J].SHS Web of Conferences, 2024, DOI:10.1051/shsconf/202418501005.