Innovative Applications and Practical Effects of Multimodal Artificial Intelligence Technology in Integrated Chinese Language Teaching

Chenyang Feng

University of International Business and Economics, Dongguan, China

1781507120@qq.com

Keywords: Multimodal AI; Integrated Chinese Course; Innovative Teaching; Practical Effects; Language Acquisition

Abstract: With the rapid development of multimodal artificial intelligence (AI), integrating text, image, audio, and interactive technologies has brought new opportunities for teaching Chinese as a second language. This paper investigates the innovative applications of multimodal AI in integrated Chinese courses and evaluates their practical effects on learners' linguistic competence and communicative ability. Through theoretical analysis and classroom practice, this study explores how multimodal AI can enhance vocabulary acquisition, discourse comprehension, and intercultural communication skills, while also addressing challenges such as over-reliance on technology and pedagogical adaptation. The findings provide both theoretical insights and practical references for promoting intelligent, personalized, and interactive Chinese language teaching.

1. Introduction

1.1. Research Background

Integrated Chinese courses, as the fundamental component of Chinese language teaching, aim to cultivate learners' comprehensive abilities in listening, speaking, reading, and writing [1]. However, in practice, traditional teacher-centered methods often encounter difficulties in maintaining students' long-term motivation, balancing the development of different skills, and addressing diverse learner needs. With the rapid advancement of educational technology, multimodal AI technologies—such as intelligent speech recognition, image–text generation, and virtual interactive platforms—have emerged as innovative tools. These technologies not only extend the range of pedagogical resources but also create immersive and interactive learning environments, thereby reshaping the teaching model of integrated Chinese courses.

1.2. Research Significance

Exploring the application of multimodal AI in integrated Chinese teaching carries both theoretical and practical significance. On the theoretical level, it enriches the interdisciplinary dialogue among applied linguistics, educational technology, and AI-assisted pedagogy, offering new perspectives on language learning mechanisms. On the practical level, it provides teachers with intelligent strategies to optimize instructional design, enhance teaching quality, and promote learner engagement. Furthermore, it helps students improve language proficiency more efficiently, while simultaneously strengthening intercultural communication competence, thus meeting the growing demand for high-quality Chinese education in global contexts.

2. Theoretical Foundation of Multimodal AI and Integrated Chinese Teaching

2.1. Concept of Multimodal AI Technology

Multimodal artificial intelligence (AI) refers to the integration and processing of information from multiple channels, including text, speech, vision, and gesture recognition [2]. Unlike traditional single-modal AI systems, multimodal AI can simultaneously analyze linguistic, auditory, and visual data, thereby enabling more dynamic human–machine interaction. For instance, intelligent speech

  DOI: 10.25236/gemmsd.2025.101

recognition allows real-time assessment of pronunciation and fluency, while image–text generation creates vivid visual contexts to support comprehension and cultural learning. Gesture and facial recognition further enhance the interactivity of virtual classrooms by enabling more natural communication. Collectively, these technologies create adaptive learning environments where learners' performance can be monitored and feedback provided instantly, thus fostering personalized and immersive language learning experiences.

2.2. Features of Integrated Chinese Teaching

Integrated Chinese teaching is a pedagogical model that focuses on the holistic cultivation of linguistic skills, aiming to develop learners' abilities in listening, speaking, reading, and writing in a balanced manner [3]. Beyond basic language training, it also emphasizes communicative competence, cultural awareness, and the ability to use language flexibly in real-life contexts. To achieve these goals, integrated courses often design diverse input channels—such as texts, dialogues, multimedia materials, and situational tasks—while encouraging learners' active participation through role play, discussions, and project-based learning. This multidimensional approach reflects the principle that language acquisition is most effective when learners are exposed to meaningful input, engaged in output activities, and provided with interactive opportunities that mirror authentic communication scenarios.

2.3. Compatibility between Multimodal AI and Integrated Teaching Goals

The pedagogical objectives of integrated Chinese courses align closely with the affordances of multimodal AI. First, AI-driven adaptive scaffolding can adjust task difficulty and provide targeted support based on learners' real-time performance, helping students at different proficiency levels progress steadily. Second, multimodal input—combining speech, text, images, and videos—offers learners an enriched linguistic and cultural environment, which supports both comprehension and production skills. Third, intelligent feedback mechanisms, such as automated error detection in pronunciation or grammar, enable learners to refine their performance continuously. Moreover, virtual interactive platforms supported by AI simulate authentic communicative contexts, giving students opportunities to practice intercultural interaction in safe and engaging ways [4]. Thus, multimodal AI not only complements but also enhances the integrated teaching model by bridging the gap between traditional instruction and personalized, technology-enhanced learning.

3. Innovative Applications of Multimodal AI in Integrated Chinese Courses

3.1. Intelligent Vocabulary and Grammar Learning

Vocabulary and grammar constitute the foundation of language competence, yet they often present difficulties for learners due to memorization fatigue and insufficient contextualization. AI-powered multimodal platforms provide effective solutions by integrating visual images, contextual dialogues, and speech synthesis to reinforce word recognition and syntactic structures [5]. For example, learners can encounter new words accompanied by pictures or short videos, which activate associative memory and facilitate deeper semantic understanding. At the same time, contextualized dialogues generated by AI demonstrate how grammatical patterns are used in authentic communicative situations. Speech synthesis further enables learners to hear accurate pronunciation and intonation, strengthening the connection between form and meaning. Such multimodal reinforcement makes vocabulary and grammar acquisition both more systematic and engaging.

3.2. Interactive Reading and Listening Comprehension

Reading and listening comprehension are essential components of integrated Chinese teaching, but traditional materials are often static and lack adaptability. Generative AI technologies can dynamically produce reading texts and audio materials tailored to learners' proficiency levels, interests, and learning objectives [6]. For instance, a beginner may receive simplified narratives with supportive glosses, whereas an advanced learner may engage with culturally rich authentic texts.

These materials can be paired with AI-generated comprehension questions that assess literal, inferential, and critical understanding. Real-time personalized feedback allows learners to identify weaknesses and improve strategies, such as skimming, scanning, or note-taking. Through this process, AI enhances both the efficiency and depth of comprehension training.

3.3. AI-Enhanced Speaking and Writing Practice

Speaking and writing are productive skills that require constant practice and timely feedback, yet in traditional classrooms, teachers may find it difficult to provide individualized support. Multimodal AI technologies address this challenge through speech recognition and natural language generation. Learners can practice oral tasks while receiving instant evaluation of pronunciation, fluency, and prosody, along with corrective feedback. Similarly, AI writing assistants can analyze learners' texts for grammar, vocabulary use, coherence, and stylistic appropriateness, offering suggestions for revision. These functions transform speaking and writing from teacher-dependent activities into autonomous practices supported by intelligent scaffolding, which promotes accuracy, confidence, and creativity.

3.4. Virtual Intercultural Communication Scenarios

Language learning is inseparable from cultural understanding, and intercultural competence has become a crucial goal of Chinese teaching in international contexts. Immersive virtual reality (VR) and augmented reality (AR) environments allow learners to participate in role-play conversations with AI-driven avatars, simulating authentic intercultural situations such as shopping in a Chinese market, attending a festival, or negotiating in a workplace setting [7]. These scenarios provide opportunities to apply linguistic knowledge in realistic contexts while also exposing learners to sociocultural norms, gestures, and pragmatic conventions. By practicing communication in a safe and engaging virtual space, learners not only strengthen language proficiency but also develop the cultural adaptability required for real-world interactions.

4. Practical Effects and Challenges of Multimodal AI Integration

4.1. Learning Outcome Enhancement

One of the most significant contributions of multimodal AI technology to integrated Chinese teaching lies in its impact on measurable learning outcomes. Empirical studies have consistently demonstrated that learners exposed to AI-supported multimodal input achieve notable improvements in several key areas. Vocabulary retention, for instance, is enhanced through the combination of image-based word presentation, contextualized examples, and repeated retrieval practices generated automatically by AI systems [8]. Listening comprehension benefits from adaptive audio materials that align with learners' proficiency levels and gradually increase in complexity. Oral fluency improves as learners engage in interactive dialogues with AI avatars or speech recognition systems, which provide immediate corrective feedback on pronunciation, rhythm, and intonation. Similarly, writing accuracy is strengthened through AI-powered text analysis tools that highlight errors in grammar, vocabulary usage, and coherence, while also offering suggestions for revision. Collectively, these outcomes indicate that multimodal AI does not merely supplement traditional teaching but actively promotes deeper learning and sustained language development.

4.2. Learner Motivation and Engagement

Beyond cognitive outcomes, multimodal AI also exerts a strong influence on affective and motivational dimensions of learning. Interactive platforms, gamified exercises, and immersive virtual scenarios transform language learning from a repetitive process into an engaging and enjoyable experience. Learners often report greater interest in participating in tasks that involve AI-mediated interactions, as these activities provide novelty, immediacy, and personalization. At the same time, AI tools can reduce learning-related anxiety, particularly in speaking tasks, by allowing students to practice with virtual partners before interacting with human peers or teachers. This safe

practice environment enhances self-confidence and encourages risk-taking, which is essential for communicative development. Moreover, the adaptive nature of AI systems ensures that learners receive appropriately challenging tasks, preventing frustration from overly difficult materials and boredom from tasks that are too simple. As a result, multimodal AI fosters sustained engagement and intrinsic motivation, which are critical for long-term language acquisition.

4.3. Pedagogical Adaptation Challenges

While the benefits of multimodal AI are evident, its integration into Chinese language classrooms also presents challenges for pedagogy and teacher development. Teachers are required to develop a certain level of AI literacy, including the ability to understand how the technology operates, how to interpret AI-generated feedback, and how to integrate digital tools into broader curricular goals [9]. This need for professional development often demands institutional support in the form of training programs and ongoing technical assistance. In addition, lesson plans must be redesigned to accommodate the interactive and adaptive nature of AI tools. Teachers must consider how to balance human guidance with AI assistance to avoid situations where technology dominates classroom interactions. For example, while AI may provide instant correction in speaking or writing, teachers must still guide students in reflecting on their errors and developing metacognitive strategies. Effective implementation thus requires teachers to play the dual role of technology mediator and pedagogical designer, ensuring that AI use remains aligned with the holistic objectives of integrated Chinese teaching.

4.4. Risks of Over-Reliance on Technology

Despite the promise of multimodal AI, there are potential risks associated with its excessive or uncritical use. One concern is the reduction of learner autonomy. When students become overly dependent on AI-generated feedback and suggestions, they may lose the ability to self-monitor and regulate their own learning processes. Another challenge involves ethical considerations surrounding data privacy and security. Since multimodal AI platforms often collect large amounts of personal learning data, issues such as consent, storage, and potential misuse require careful attention from educators and institutions. Furthermore, there is the danger of superficial engagement, in which learners interact mechanically with AI tools without developing deeper critical thinking or reflective skills. For instance, students may rely on automatic corrections in writing tasks without analyzing why their output was incorrect, leading to short-term improvements but limited long-term gains. Therefore, the successful application of multimodal AI requires a balanced approach: technology should be harnessed to enhance, rather than replace, human agency, critical reflection, and meaningful interaction in the language learning process.

5. Evaluation and Future Prospects

5.1. Evaluation Indicators

Evaluating the effectiveness of multimodal AI integration in Chinese language teaching requires a strategically designed framework that captures both cognitive and affective dimensions of learning outcomes. From a strategic perspective, key indicators should not only measure immediate performance gains but also guide ongoing instructional decisions. Learning achievement remains a core metric, encompassing measurable improvements in vocabulary acquisition, grammatical accuracy, reading comprehension, listening comprehension, oral fluency, and writing proficiency. Communicative competence is equally critical, as it reflects learners' ability to apply language knowledge in authentic interpersonal and intercultural contexts, which is central to integrated teaching objectives [10]. Learner satisfaction should be monitored strategically, focusing on engagement, motivation, confidence, and perceived value of AI-assisted tasks, since these factors influence persistence and self-directed learning. Teacher adaptability constitutes another essential indicator, highlighting instructors' capacity to integrate AI tools effectively, design adaptive lesson plans, and maintain a balanced interplay between human guidance and AI support. Strategically

tracking these indicators provides actionable insights for optimizing teaching practices and ensuring sustainable implementation of AI-enhanced pedagogy.

5.2. Evaluation Methods

A mixed-methods strategy is particularly effective for capturing the multifaceted impact of multimodal AI in integrated Chinese courses. Quantitative methods, such as pre- and post-tests, measure objective improvements in learners' language proficiency, while AI-generated analytics offer real-time tracking of engagement patterns, error types, and learning trajectories. These data can inform targeted interventions, such as adjusting task difficulty, scaffolding content, or providing additional practice for weak areas. Qualitative methods—classroom observations, semi-structured interviews with students and teachers, and reflective journals—offer nuanced insights into interaction patterns, classroom dynamics, and behavioral shifts. Learner surveys, incorporating both Likert-scale questions and open-ended reflections, provide additional evidence regarding usability, satisfaction, and motivational impact. By strategically triangulating these data sources, educators and researchers can obtain a comprehensive understanding of AI-enhanced teaching outcomes and generate evidence-based recommendations for instructional improvement. Furthermore, the longitudinal use of AI analytics supports continuous monitoring, enabling educators to identify emerging trends, anticipate challenges, and iteratively refine teaching strategies to maximize learning effectiveness.

5.3. Future Prospects and Strategic Implications

Looking forward, the strategic application of multimodal AI in Chinese language education should emphasize intelligent personalization, adaptive guidance, and the creation of integrated learning ecosystems. AI-powered tutoring systems can deliver individualized learning pathways, dynamically adjusting content, pacing, and difficulty to each learner's profile, thereby supporting differentiated instruction. Cross-platform integration will enable seamless connections between classroom teaching, online resources, mobile applications, and immersive virtual environments, ensuring continuity of learning beyond traditional classroom boundaries. Strategically designed intelligent learning ecosystems—combining speech recognition, natural language generation, adaptive feedback, and VR/AR-based cultural simulations—can provide highly interactive, contextually rich, and culturally immersive experiences [11]. These ecosystems may also incorporate collaborative and social learning features, allowing students to communicate with peers, engage with native speakers, and participate in global language communities, thereby enhancing both linguistic and intercultural competence.

From a pedagogical strategy perspective, AI technologies should also support teachers in curriculum design, formative assessment, and instructional decision-making. By automating routine feedback, tracking learner progress, and providing actionable insights, AI allows educators to focus on higher-order teaching tasks, such as facilitating critical thinking, promoting reflective learning, and guiding authentic communication activities. Importantly, future strategies must address ethical and sustainable integration, ensuring responsible data use, protecting learner privacy, and balancing AI reliance with human-centered teaching. Overall, the strategic combination of intelligent personalization, cross-platform accessibility, immersive interaction, and data-informed instructional planning promises to transform Chinese language education. It enables not only more efficient and engaging learning but also the cultivation of culturally competent, autonomous, and lifelong learners in the era of AI-driven education.

6. Conclusion

This study explored the innovative applications and practical effects of multimodal AI technology in integrated Chinese language teaching. The findings indicate that the integration of multimodal AI significantly enhances learners' linguistic competence across listening, speaking, reading, and writing, strengthens their motivation and engagement, and facilitates the development of intercultural communication skills in authentic and immersive contexts. At the same time, the

study highlights several challenges, including the need for pedagogical adaptation, the cultivation of teachers' AI literacy, and the careful management of potential risks such as over-reliance on technology and ethical concerns related to data use. Therefore, future development in this area should emphasize effective teacher–AI collaboration, responsible and transparent use of learner data, and the establishment of sustainable, human-centered integration models. Such strategies will help ensure that multimodal AI not only enriches instructional practices but also maintains balanced, high-quality, and culturally meaningful Chinese language teaching in the era of advanced educational technologies.

References

[1] XIONG Y. The integrated application of Chinese Baijiu culture and storytelling in teaching Chinese as a second language[J]. SHS Web of Conferences, 2024, 190:1025. DOI:10.1051/shsconf/202419001025.

[2] Dicker A P , Ma X , Tang S ,et al.Impact of specimen type on digital histopathology-based multimodal artificial intelligence (MMAI) biomarker risk score: Whole slide image vs tissue microarray.[J].Journal of Clinical Oncology, 2025, 43(5_suppl):400.DOI:10.1200/jco.2025.43.5_suppl.400.

[3] Yang C , Wei X , Xue L .Application and Effectiveness of Task-Based Language Teaching (TBLT) in Teaching Chinese as a Second Language[J].Advances in Humanities Research, 2025, 11.DOI:10.54254/2753-7080/2025.20563.

[4] Phuong N T T , Nhung T T H , Hiep V T N .Teaching integrated with cultural and social contexts for sustainable development goals[J].International Journal of Research Studies in Education, 2025, 14(3).DOI:10.5861/ijrse.2025.25822.

[5] Kovari A .A systematic review of AI-powered collaborative learning in higher education: Trends and outcomes from the last decade[J].Social Sciences & Humanities Open, 2025, 11.DOI:10.1016/j.ssaho.2025.101335.

[6] Yeh H C .Revolutionizing language learning: Integrating generative AI for enhanced language proficiency[J].Educational Technology & Society, 2024, 27(3).DOI:10.30191/ETS.202407_27(3).TP01.

[7] Rahmad M , Arafah W , Ingkadijaya R ,et al.Utilization of Immersive Technology in Qualitative Research Design: Virtual Reality (VR) and Augmented Reality (AR) Opportunities[J].Jurnal Indonesia Sosial Teknologi, 2024, 5(12).DOI:10.59141/jist.v5i12.8816.

[8] Abdel Latif O L I .Using Infographics to Enhance STEM Students' Vocabulary Retention and Visual Literacy[J].Journal of Research in Curriculum, Instruction & Educational Technology, 2024, 10(2).DOI:10.21608/jrciet.2024.363669.

[9] Karpovich I , Voronova L , Ivanova T ,et al.Challenges of Professional Adaptation: Difference Between Novice University Teachers with and Without Pedagogical Training[C]//International Conference on Professional Culture of the Specialist of the Future.Springer, Cham, 2022. DOI:10.1007/978-3-030-89708-6_58.

[10] Young R F .Interactional competence: Challenges for validity[J].CiteSeer, 2000.

[11] Kong Y , Zhang M , Ye D ,et al.An intelligent agent-based method for task allocation in competitive cloud environments[J].Concurrency & Computation Practice & Experience, 2017: e4178.DOI:10.1002/cpe.4178.