Research on the Application Scenarios and Limitations of Artificial Intelligence-Assisted Chinese Oral Language Teaching

Yanling Chen[*]

HeXi University, Zhangye, 734000, China
[*]Corresponding author: yanl315@126.com

Abstract: *With the breakthrough advancements in artificial intelligence technology, its deep integration with second language acquisition theories is reshaping the structure and paradigms of Chinese oral language teaching. This study systematically explores the theoretical mechanisms, multidimensional applications, and inherent limitations of AI-assisted Chinese oral language instruction. By analyzing the transformation of teaching paradigms and the deconstruction of competency components driven by intelligent technology, the research constructs systematic application scenarios across dimensions such as pronunciation feedback, contextual dialogue, adaptive learning, and immersive environments. Simultaneously, the study reveals multiple constraints currently faced by technology-enabled teaching, including technical bottlenecks in cultivating complex communicative competence, adaptive limitations in teaching Chinese tones and pragmatic cultural elements, fairness concerns arising from data and algorithms, and ethical challenges posed by the blurring boundaries between humans and machines. This study aims to provide a systematic analytical framework for prudently advancing the deep integration of artificial intelligence and Chinese oral language teaching.*

Keywords: *Artificial Intelligence; Chinese Oral Language Teaching; Application Scenarios; Teaching Paradigm; Technological Limitations; Human-Computer Interaction*

Introduction

The core of Chinese oral language teaching lies in cultivating learners' ability to communicate accurately, fluently, and appropriately in authentic contexts. However, traditional models are constrained by resource allocation, making it difficult to meet individual needs for massive input, instant feedback, and personalized learning pathways. Artificial intelligence technologies, represented by deep learning and natural language processing, provide new pathways to address these challenges. Their significance extends beyond mere tool application; from the intersection of cognitive and linguistic theories, they drive the restructuring of teaching frameworks, competency composition, and evaluation methods. Existing research tends either towards optimistic technological discourse or is confined to the analysis of single tools, lacking systematic integration of application scenarios, theoretical foundations, and their inherent limitations. Therefore, analyzing the theoretical coupling mechanism between artificial intelligence and Chinese oral language teaching, constructing multidimensional application scenarios, and reflecting on its adaptive boundaries and ethical challenges hold significant theoretical value and practical necessity. This study aims to transcend a mere instrumental view of technology and delve deeply into the structural transformations and fundamental constraints brought by intelligentization to Chinese oral language teaching.

1. Theoretical Convergence and Mechanism Analysis of Artificial Intelligence and Chinese Oral Language Teaching

1.1 Cognitive Theoretical Foundations of Second Language Oral Acquisition

The construction of second language oral proficiency is rooted in complex cognitive psychological processes. Among these, the Input Hypothesis and the Interaction Hypothesis emphasize the importance of comprehensible language input and meaning negotiation for language internalization. The Output Hypothesis posits that learners must engage in compulsory language production to drive their transition from semantic to syntactic processing, thereby enhancing linguistic accuracy and automaticity. Information Processing Theory views language learning as the allocation of limited cognitive resources among attention, memory, and processing, with the development of oral fluency relying on the formation of automated processing to reduce cognitive load. Sociocultural Theory further regards dialogue as the medium for the development of thinking and language, emphasizing the construction of knowledge and skills through collaborative dialogue. These theories collectively outline the key features of an ideal oral language teaching environment: providing ample comprehensible input that is slightly beyond the current proficiency level, creating frequent and cognitively challenging opportunities for meaning negotiation and production, and fostering supportive interactions to promote internalization. This provides a clear theoretical direction for the intervention of artificial intelligence and establishes a cognitive dimension as a benchmark for evaluating its effectiveness[1].

1.2 Transformation of Teaching Models Driven by Core Artificial Intelligence Technologies

Contemporary artificial intelligence technologies, represented by deep learning, are penetrating beyond the instrumental level to the core of pedagogical structure, driving a structural transformation in Chinese oral language teaching models. Intelligent speech processing technologies, particularly automatic speech recognition and speech synthesis, enable machines to process learners' speech signals with unprecedented precision and real-time capability, achieving quantitative analysis and imitative feedback ranging from pronunciation to prosody. Natural language processing technologies, including syntactic analysis, semantic understanding, and dialogue generation, allow systems to parse the structure and meaning of learners' oral output and generate dynamic contexts and interactive content aligned with specific instructional objectives. Computer vision and multimodal fusion technologies further expand the dimensions of interaction by enabling the recognition and response to learners' paralinguistic information, such as facial expressions and gestures, thereby creating more natural and multidimensional communicative scenarios. The integrated application of these technologies promotes a shift in teaching models from the traditional linear transmission-characterized by teacher-centeredness, fixed content, and uniform pacing-towards a personalized and intelligent developmental pathway that is based on individual learners' data streams and features dynamic adaptation and high interactivity.

1.3 Deconstruction of Oral Proficiency Components Enabled by Intelligent Technology

Empowered by intelligent technology, the components of oral proficiency, traditionally regarded as an integrated whole, can be deconstructed, quantified, and targeted for intervention. Pronunciation ability can be broken down into discrete feature points such as initials, finals, tones, intonation, and rhythm. Speech recognition algorithms enable millisecond-level positioning and deviation diagnosis of these features, providing visualizable, comparable, and precise corrective guidance. Fluency indicators, including speech rate, pause frequency and duration, repetitions, and self-corrections, can be automatically extracted and analyzed by algorithms, offering data support for enhancing the automation and coherence of oral output. The ability to use vocabulary and grammar can be assessed through natural language processing technologies, which analyze the diversity and complexity of vocabulary as well as the accuracy of syntactic structures in learners' real-time output, identifying cross-linguistic transfer effects and interlanguage development characteristics. The most core aspect, communicative competence-particularly pragmatic appropriateness and cultural adaptability-though more complex, can also be addressed by constructing rich, culturally annotated dialogue scenarios and multi-turn human-computer interactions, providing learners with opportunities to experiment, receive feedback, and adjust within a controlled risk environment. This fine-grained deconstruction at the component level serves as the technical prerequisite for achieving precise instruction and personalized development[2].

1.4 Reconstruction Paths for Oral Language Teaching Paradigms in the Context of Human-Computer Interaction

The deep integration of artificial intelligence has given rise to a novel "human-computer collaboration" teaching paradigm. Its reconstruction paths are primarily reflected across three dimensions: the relationship between teaching agents, the logic of the teaching process, and the attributes of the teaching environment. At the level of teaching agents, the teacher's role is evolving from that of a sole knowledge transmitter and authoritative evaluator into a designer of learning environments, a coordinator of human-computer collaboration, and a guide for emotional and higher-order cognitive development. Meanwhile, artificial intelligence assumes the duties of a tireless practice partner, an instant feedback provider, and an engine for personalized content. Regarding the logic of the teaching process, instruction is shifting from a predetermined, fixed script to a dynamically generated script driven by real-time data, where learning paths are adaptively adjusted and optimized based on the learner's performance data. In terms of the teaching environment's attributes, intelligent technology constructs an "extended learning environment." This environment can serve as a highly structured training ground for specific skills or as an open-ended sandbox simulating authentic communication, blending the virtual and the real to transcend limitations of time, space, and physical resources. The core of this paradigm lies in organically combining the machine's advantages of standardization and data-driven capabilities with the human teacher's flexibility, creativity, and emotional intelligence, jointly acting upon the complex ecosystem of the learner's oral communicative competence.

2. Multi-dimensional Application Scenario Construction of Artificial Intelligence in Chinese Oral Language Teaching

2.1 Real-time Feedback Mechanisms of Intelligent Speech Recognition and Pronunciation Correction Systems

Intelligent speech recognition technology, based on deep neural networks, provides an immediate feedback mechanism for Chinese pronunciation training that surpasses traditional methods. This system, trained on vast databases of Chinese speech, constructs sophisticated acoustic and language models, enabling high-accuracy recognition and transcription of learner speech input. Its core application lies in the fine-grained diagnosis of Chinese phonetic elements. The system can automatically detect and locate mispronunciations of initials and finals, particularly distinguishing between Chinese-specific features such as aspirated and unaspirated sounds, as well as retroflex and apical sounds. At the tonal level, algorithms can extract pitch contour curves, compare them with standard models of the four tones and the neutral tone, and visually present deviations in tone shape and pitch value. This feedback is highly immediate and individually targeted, providing objective, quantifiable data reports and comparative audio samples within moments of learner production, thereby guiding learners through cycles of self-monitoring and correction. This mechanism transforms pronunciation correction-which traditionally relied on teachers' subjective auditory judgment and limited classroom time-into a standardized, data-driven training process accessible at any time, effectively supporting the development of automated pronunciation skills.

2.2 Contextualized Dialogue Simulation and Generation Driven by Natural Language Processing

Natural language processing technology makes it possible to create dynamic and rich contextualized dialogue simulations, directly serving the cultivation of oral communicative competence. Such systems, relying on dialogue state tracking, natural language understanding, and natural language generation technologies, can construct virtual dialogue environments containing specific communicative goals, social roles, and cultural scenarios. Based on the learner's language proficiency and teaching topics, the system can dynamically generate dialogue prompts and responses, guiding the conversation to progress over multiple turns. Its key advantages lie in the diversity and controllability of contexts, allowing learners to engage in low-risk repetitive practice in simulated scenarios such as shopping, asking for directions, social interactions, and debates. Furthermore, advanced systems can identify pragmatic errors made by learners during dialogues-such as inappropriate forms of address, misjudgment of politeness levels, or breaches of cultural taboos-and provide alternative expressions that conform to the cultural norms of the target language. This form of dialogue simulation driven by natural language processing not only trains the accuracy of linguistic forms but also situates language within its sociocultural framework of use, providing structured support for developing learners' contextual adaptability and communicative strategies[3].

2.3 Adaptive Learning Pathways and Content Delivery Based on Learner Profiles

Artificial intelligence enables the construction of dynamically evolving cognitive profiles of learners through the continuous collection and analysis of multi-dimensional learning data, which in turn drives the operation of adaptive learning systems. These profiles integrate multi-source data, including the learner's pronunciation error patterns, frequently used vocabulary and syntactic structures, dialogue fluency metrics, practice frequency and duration, as well as success rates and challenging points in specific tasks. Based on this information, machine learning algorithms identify the learner's areas of strength, weaknesses, and personalized learning patterns. The system then utilizes these insights to dynamically adjust the difficulty and sequence of the learning pathway-for instance, providing targeted comparative training for learners who struggle with tone confusion, or reinforcing semantic field exercises for those facing difficulties with vocabulary retrieval. Content delivery is also personalized, filtering and generating dialogue materials and expression examples from vast corpora that match the learner's current language proficiency, acquired knowledge, and potential interests. This scenario achieves a shift from a standardized curriculum to "one path per individual," ensuring that the allocation of teaching resources consistently revolves around the optimal zone of development for each learner, thereby enhancing learning efficiency.

2.4 Creation of Immersive Oral Communication Environments through Multimodal Fusion

Multimodal systems integrating technologies such as virtual reality, augmented reality, computer vision, and spatial audio can create highly immersive Chinese oral communication environments that approximate real-world language use experiences. In such environments, learners are placed into panoramic Chinese sociocultural scenarios-such as virtual traditional courtyards, marketplaces, or meeting rooms-through head-mounted devices or interactive interfaces. The system not only provides visual and auditory immersion but also captures learners' body movements, gaze direction, and spatial positioning through sensors, enabling real-time interaction with elements in the virtual environment. For example, learners may need to collaborate with virtual characters through oral instructions to complete tasks or produce appropriate verbal responses based on environmental cues. Spatial audio technology ensures the directionality and sense of distance of sounds, enhancing the authenticity of communication. This multimodal immersive environment deeply integrates linguistic forms, paralinguistic information, cultural context, and bodily experience, powerfully promoting the integrative acquisition of language knowledge and communicative skills. It provides an advanced training ground for cultivating the ability to engage in effective Chinese communication in complex, dynamic real-world contexts[4].

3. Real-world Limitations and Profound Challenges of AI-assisted Chinese Oral Language Teaching

3.1 Constraints of Technological Bottlenecks on Cultivating Complex Oral Communicative Competence

Although artificial intelligence demonstrates significant capabilities in speech recognition and structured dialogue, it still faces fundamental technological bottlenecks in cultivating complex oral communicative competence that involves higher-order cognition and social interaction. Current systems possess limited capacity for processing paralinguistic information; they struggle to accurately interpret and generate the emotional attitudes and communicative intentions conveyed by the facial expressions, gestures, posture, and prosodic variations accompanying spoken interaction. At the level of contextual understanding, algorithms are deficient in their ability to model and reason in real-time about dynamic, shared, and implicit contextual information. This results in systems finding it difficult to handle situational meanings, humor, irony, or culturally specific expressions that rely on concrete contexts. For complex communicative tasks requiring the immediate, creative organization of language and multi-turn negotiation of viewpoints and logical argumentation-such as in academic discussions or business negotiations-the outputs of generative models often exhibit limitations in logical coherence, depth of perspective, and strategic flexibility. These bottlenecks mean that current technology-assisted oral training predominantly focuses on pre-defined, well-structured micro-skills, while support for integrated, adaptive, and creative macro-communicative competence remains weak. This constitutes a key obstacle in transitioning from "standardized training" to "intelligent enablement".

3.2 Adaptability Limits of Intelligent Systems in Teaching Chinese Tones and Pragmatic-Cultural Elements

The unique linguistic and cultural attributes of the Chinese language pose distinct challenges to artificial intelligence systems, revealing their limitations in facilitating deep language acquisition. At the phonetic level, the Chinese tonal system involves not only static pitch targets but is also intricately interwoven with coarticulation in connected speech, intonation, and emotional expression, forming dynamic and relative tonal patterns. Most existing speech recognition and evaluation models are trained on isolated syllables or standardized sentences; consequently, their analysis and feedback regarding the flexible variations of tones in continuous speech, as well as the complex interaction between neutral tones and tone sandhi rules, lack sufficient accuracy and pedagogical explanatory power. At the pragmatic and cultural level, Chinese communication is highly dependent on context, relational hierarchies, and cultural norms, such as politeness strategies, forms of address, and expressions of humility and respect. While current models trained on large-scale text corpora can identify some superficial forms, they struggle to deeply understand and simulate the dynamic, nuanced cultural-cognitive frameworks and social-situational judgments underlying these norms. Systems may generate grammatically correct but pragmatically inappropriate sentences, or provide overly simplified or unrealistic explanations in teaching scenarios rich with cultural connotations. This may lead learners to grasp the form while neglecting appropriateness, posing the risk of developing an interlanguage that is "fluent yet inappropriate"[5].

3.3 Questions of Instructional Equity Arising from Data Dependency and Algorithmic Bias

The efficacy of artificial intelligence systems is highly dependent on the scale, quality, and representativeness of their training data. This data dependency harbors the risk of precipitating issues of instructional equity. If the datasets used to train speech recognition and evaluation models lack sufficient coverage in terms of speaker age, regional dialect, gender, and linguistic background, it may lead to decreased speech recognition accuracy and unfair feedback quality for specific learner groups, such as those with particular native language backgrounds or regional accents. At the level of natural language processing and content generation, latent biases present in the training corpora-pertaining to cultural perspectives, value orientations, or discourse styles-may be amplified and entrenched by algorithms within the generated instructional materials and interactive responses. This could inadvertently convey singular or biased cultural representations and worldviews to learners. Furthermore, the pathway planning and resource allocation conducted by adaptive learning systems based on learners' historical data, if poorly designed, may fall into the trap of "filter bubbles" or the "Matthew Effect." This could restrict learners' exposure to diverse language variations and cultural content, or, for learners in disadvantaged positions, persistently assign low-challenge tasks due to poor initial data, thereby exacerbating disparities in competency development. These questions of equity, arising from the inherent characteristics of data and algorithms, constitute ethical issues that must be carefully considered when applying technology to inclusive education.

3.4 Reflections on Teaching Ethics Arising from the Blurring of Human-Computer Interaction Boundaries

The deepening role of artificial intelligence in oral language teaching and the fusion of human-computer interfaces have prompted an expansion and reconsideration of the scope of teaching ethics. The vast amounts of learner behavioral and speech data generated during the teaching process raise profound issues of privacy protection and data ethics, involving questions of data ownership, storage security, access permissions, and anonymization. When systems possess affective computing capabilities, attempting to recognize and respond to learners' emotional states, they touch upon the boundaries of emotional privacy and raise concerns about the legitimacy and risks of algorithmically intervening in a learner's psychological state. At the level of instructional design, excessive reliance on or absolute trust in the assessments and decisions of intelligent systems may undermine the critical, leading role of teaching professionals, reducing the complex educational process to a mere technical optimization problem. The division of responsibility within human-computer collaboration becomes blurred; for instance, accountability mechanisms remain undefined for situations where systems provide erroneous feedback or inappropriate content that hinders learning. A deeper ethical reflection concerns whether, and to what extent, highly personalized intelligent systems-which shape learners' language habits and cognitive pathways-should be regarded as an implicit cultural-cognitive mediator. The long-term impact of the value assumptions and technological rationality embedded in their design on learners' linguistic identity construction and cultural recognition urgently demands interdisciplinary attention and prospective discussion.

Conclusion

This study systematically explores the theoretical foundations, multidimensional applications, and inherent limitations of artificial intelligence-assisted Chinese oral language teaching. The research finds that the deep integration of intelligent technology and second language acquisition theory is driving a transformation towards a data-driven, personalized, and intelligent pedagogical paradigm, thereby reshaping the relationships between teaching agents and the attributes of the learning environment. Application scenarios such as intelligent speech feedback, contextual dialogue simulation, adaptive learning, and multimodal immersion demonstrate significant potential in pronunciation training, communicative competence cultivation, and higher-order skill integration. However, technological enablement also reveals its intrinsic boundaries: algorithms have limited capacity for processing paralinguistic information and supporting dynamic contexts in complex communication; their adaptability to the dynamic nature of Chinese tones and the deep-seated pragmatic-cultural aspects of the language remains insufficient; data and algorithms harbor latent fairness risks; and the deepening of human-computer collaboration raises issues concerning data ethics and accountability. Future research needs to break through the technical bottlenecks in complex context modeling and affective interaction, develop algorithmic models with greater cultural adaptability for Chinese, construct inclusive datasets and evaluation frameworks, and deepen the exploration of teaching ethics within human-computer collaboration. The ultimate value of artificial intelligence lies in constructing a new educational ecology that stimulates teacher wisdom and learner potential through prudent integration.

References

[1] Zhang Yue. "Instructional Design for Digital-Intelligent Project-Based Learning Supported by Generative Artificial Intelligence." 2025. Shanghai International Studies University, MA thesis.

[2] Chen Yue. "Research on the Application Prospects of Artificial Intelligence in International Chinese Education." 2024. Xi’ an Shiyou University, MA thesis.

[3] Chen Si. "Research on the Promotion and Application of Artificial Intelligence in International Chinese Education." 2024. Bohai University, MA thesis.

[4] Li Jiatao. "Research on the Application of Generative Artificial Intelligence in Intermediate Chinese Oral Courses." 2025. Xi’ an Shiyou University, MA thesis.

[5] Xiao Shijun. "Phased Application of Artificial Intelligence in Chinese Oral Testing." Digital International Chinese Education (2022). Ed. School of International Studies, Macau University of Science and Technology, 2022, pp. 411-418.