



Exploring the Interactive Application of Artificial Intelligence in Oral English Teaching

Cheng Sheng

Hubei Preschool Teachers College, Wuhan 430223, Hubei, China.

How to cite this paper: Cheng Sheng. (2025). Exploring the Interactive Application of Artificial Intelligence in Oral English Teaching. *The Educational Review, USA*, 9(5), 502-507.
DOI: 10.26855/er.2025.05.004

Received: March 31, 2025

Accepted: April 30, 2025

Published: May 29, 2025

Corresponding author: Cheng Sheng, Hubei Preschool Teachers College, Wuhan 430223, Hubei, China.

Abstract

With the increasing development of digital education, the demand for interactivity and personalization in English oral teaching is becoming more urgent. The research focuses on the interactive application of artificial intelligence technology in oral teaching, aiming to explore how it can meet these needs. The application of various artificial intelligence technology models in oral teaching was studied and analyzed, including real-time feedback of speech recognition, personalized guidance of dialogue systems, simulation of virtual character scenarios, data-driven path optimization, and multimodal human-machine collaboration. Through the deconstruction and practical verification of technical functions, the effectiveness of these technologies in actual teaching has been evaluated. The study focuses on English oral learners, especially those who hope to receive more support in terms of interactivity and personalization. Research has found that artificial intelligence technology can significantly improve practice frequency, optimize feedback efficiency, and enhance situational realism in English oral teaching. These technologies break through the limitations of traditional interactive teaching and provide learners with high-frequency and precise interactive experiences. This study has important guiding significance for the innovation of smart language education and provides new directions for the future development of English oral teaching.

Keywords

Artificial Intelligence; Oral English Teaching; Interactivity; Speech Recognition; Virtual Characters

Introduction

In the wave of globalization and educational digitization, the interactive effectiveness of oral English teaching has become a key bottleneck restricting the improvement of language application skills. In traditional models, issues such as the temporal and spatial limitations of teacher-student interaction, delayed feedback, and single scenarios urgently need to be addressed. Artificial intelligence technology, with its capabilities in speech recognition and natural language processing, offers the possibility of constructing a new interactive ecosystem characterized by "real-time feedback - personalized guidance - immersive experience." Exploring the mechanisms and pathways of its interactive application in oral English teaching not only responds to the development needs of smart education but also provides an innovative paradigm for the transformation of language teaching from "one-way transmission" to "intelligent interaction," possessing dual values of theoretical breakthrough and practical optimization.

1. Interactive Voice Recognition

1.1 Real-time Pronunciation Feedback

Voice recognition technology, through the collaborative operation of acoustic and language models, constructs an instant interactive link for oral training. The system converts learners' voice signals into digital phoneme sequences and dynamically compares them with a standard pronunciation database, accurately identifying the timing and type of pronunciation deviations. For example, in cases of confusion between the "th" sounds (/θ/ and /ð/), the system can recognize pronunciation errors within 0.3 seconds, simultaneously tagging differences in parameters such as mouth opening and tongue tip position, and presenting the error features in the form of waveform graphs overlaid with textual annotations. This feedback mechanism breaks through the efficiency bottleneck of traditional teacher-led sentence-by-sentence correction, enabling learners to instantly obtain a closed-loop interactive experience of "pronunciation-evaluation-correction" during follow-up practice, increasing the effective error correction frequency of a single training session to 5-8 times that of traditional models.

1.2 Accent Correction Mechanism

Taking the common "r/l" confusion among Chinese native speakers as an example, the system analyzes comparative data between Mandarin Chinese and the English sound system, establishing an accent recognition model with 23 characteristic dimensions, which can accurately identify pronunciation deviations in easily confused words such as "light" and "right" (Tong & Liu, 2025). During the correction phase, the system combines speech organ movement simulation technology to demonstrate the trajectory of tongue and lip movements for the target phonemes in the form of three-dimensional animations, while also providing interactive functions such as slow playback and segmented practice. Experimental data show that after 12 weeks of targeted training, learners' accuracy in pronouncing the target phonemes can increase by 41.7%, and the interference of accent on listening comprehension can be reduced by 35.2%.

2. Dialogue System Interaction

2.1 Personalized Question Chains

The dialogue system, based on natural language processing, semantic understanding, and dialogue management techniques, constructs a hierarchical question generation mechanism. The system dynamically generates question sequences that match the learner's language level by analyzing the vocabulary complexity, sentence structure, and semantic coherence of the learner's input text. At the basic stage, for simple responses (such as "Yes, I do"), the system triggers supplementary questions like "Can you elaborate on why you prefer that?" to guide learners to expand their expressions; at the advanced stage, if complex sentences (such as those containing attributive clauses) are identified, critical questions such as "What counterarguments might exist for this perspective?" are further proposed to promote deeper thinking (Wei, 2024). This question chain mechanism breaks through the limitations of traditional fixed question banks, achieving a dynamic interactive loop of "input analysis—ability assessment—question generation." Experiments show that learners using this system increase the average number of effective words produced per round of dialogue by 2.3 times within 12 weeks, with a 58% increase in the use of complex sentences.

2.2 Dynamic Grammar Correction

The grammar correction function, through syntactic analysis and error pattern matching techniques, realizes real-time detection and implicit guidance of grammatical issues in oral expressions. After receiving voice input, the system first performs part-of-speech tagging and dependency syntax analysis to construct the grammatical tree structure of the sentence, and then matches it with an internal error rule library (containing 18 common error patterns such as tense, voice, and subject-verb agreement) (Pitychoutis & Rawahi, 2024). When an error is detected (such as "He go to school"), the system adopts a "semantic preservation-form correction" strategy, providing feedback in a natural conversational manner: "I notice someone went to school. Could you tell me when he did that?" This approach avoids the authoritative pressure of directly pointing out errors, instead guiding learners to independently

discover grammatical issues through context reconstruction. This mode increases learners' self-correction rate of errors to 73%, significantly higher than the 41% of traditional direct correction methods. The system automatically generates personal grammar error heatmaps, marking high-frequency error types (such as a learner making 27 instances of "third-person singular +s" errors), and pushes specialized grammar micro-exercises (such as context-prompted fill-in-the-blank training), forming a closed-loop interaction of "identification—guidance—reinforcement."

3. Virtual Character Conversation

3.1 Immersive Scenario Simulation

Virtual character technology, through multimodal scene modeling, constructs highly realistic language application environments. The system designs interactive scripts based on real-life conversation scenarios (such as business negotiations, classroom presentations, travel communication, etc.), combining the expression animations, body movements, and environmental sounds of virtual characters to create an immersive experience that integrates "visual-auditory-language" senses. In business negotiation simulations, virtual negotiators adjust their tone according to the learner's pricing strategies (such as seriously inquiring about details or smiling to indicate compromise), while dynamically presenting conference room scenes and data charts in the background to enhance situational authenticity (Cai, 2024). This simulation mechanism breaks through the time and space limitations of traditional role-playing, allowing learners to repeatedly practice complex conversation processes (such as inquiry-counteroffer-deal) in a safe virtual environment, with a single training session covering 8-12 detailed scenario nodes. Research shows that learners using this technology have a 49% improvement in language fluency in real-life situations, with a key information transmission accuracy rate of 87%, significantly higher than the 63% of traditional situational teaching methods. The system supports custom scenario functions, allowing learners to upload specific scenario images and texts to generate personalized simulation scripts, achieving an interactive upgrade from "preset scenarios" to "autonomous construction."

3.2 Emotional Interaction Design

Virtual characters, through emotional computing and dynamic response technology, build warm language interaction experiences. The system is equipped with facial expression recognition and voice emotion analysis modules, which can real-time capture learners' anxiety, confusion, and other emotional signals (such as increased frequency of frowning, faster speech speed), and trigger corresponding emotional feedback mechanisms: When detecting signs of tension, virtual characters adjust their tone and expressions, using encouraging language (such as "Take your time, this is a safe space to practice") to relieve stress; if progress signals are identified (such as correctly expressing complex viewpoints three times in a row), they provide affirmation through nonverbal cues like nodding and smiling (Wang, Zhang, & Chen, 2025). This emotional interaction breaks the mechanical nature of traditional AI interactions, making learners feel a supportive atmosphere of being understood. Experimental data indicates that the emotional design increases learners' willingness to speak by 62%, with the average duration of a single conversation extending from 8 minutes to 17 minutes. Virtual characters build personalized motivation models based on learners' emotional data, adding progressive task unlocking mechanisms for introverted learners and designing competitive challenge modes for extroverted ones, achieving precise alignment of emotional needs and learning styles.

4. Data-driven Interaction

4.1 Competence Profiling

Competence profiling achieves a three-dimensional modeling of learners' oral abilities through multi-source data integration and machine learning algorithms. The system collects pronunciation feature data generated by voice recognition (such as phoneme accuracy, intonation fluctuation range), language usage data recorded by the dialogue system (such as vocabulary diversity, grammatical complexity), and situational performance data from virtual character interactions (such as depth of topic expansion, reasonableness of emotional responses), constructing an evaluation system that includes four dimensions—phonetics, vocabulary, grammar, pragmatics—and 12 core indicators (Zhang, 2024). In the phonetic dimension, dynamic time warping (DTW) algorithm is used to analyze pronunciation

stability, quantifying parameters like "pitch deviation" and "rhythm balance"; in the pragmatic dimension, based on dialogue act classification models (such as requests, explanations, refutations), the effectiveness of communication strategies is assessed. This multidimensional data collection mechanism breaks through the limitations of traditional teaching that relies solely on teachers' subjective evaluations, reducing the error rate of competence assessment to 9.2%. The core feature of the competence profile is its dynamic updating mechanism. The system performs rolling analysis of data on a weekly basis, identifying competence development trends and potential issues by comparing learners' current performance with historical data. By introducing clustering analysis algorithms, individual data is benchmarked against group data of learners at similar levels, generating "competence radar charts" and "gap analysis reports." These clearly present learners' strengths and weaknesses in dimensions such as pronunciation accuracy and contextual adaptability. For example, a learner's radar chart shows their "vocabulary richness" is 15% higher than the group average, but their "cross-cultural expression appropriateness" is 22% lower than the average. Based on this, the system determines that they need to strengthen pragmatic training in specific cultural contexts.

4.2 Learning Path Optimization

Learning path optimization, grounded in competence profiles, generates dynamically adaptive training programs through reinforcement learning algorithms. The system first constructs a path framework consisting of three stages—basic consolidation, advanced enhancement, and extended application—based on the learner's current ability level and target differences. In the basic stage, targeted micro-skill training (such as phoneme decomposition exercises for those with poor pronunciation) is prioritized; in the advanced stage, cross-skill comprehensive tasks (such as combining grammar correction with scenario dialogues to train accurate expression in complex situations) are designed using situational simulation technology; in the extended stage, real-world corpora (such as TED talks, academic discussions) are introduced for critical expression training, promoting language ability to higher levels (Jiang et al., 2022). The dynamic adjustment mechanism of personalized paths is reflected in three aspects: First, a real-time response mechanism where the system automatically reduces difficulty and adds guiding prompts (such as keyword fill-in-the-blank assistance in dialogue practice) when a learner's accuracy falls below 70% for three consecutive times in a segment; second, a preference adaptation mechanism that analyzes learners' interaction behavior data (such as a tendency to choose workplace or daily scenarios) to intelligently adjust the proportion of training content—for example, a working professional's training program has a 65% focus on "business negotiation" scenarios, significantly higher than the default 30%; third, a goal-oriented mechanism that allows learners to set short-term goals (such as improving interview oral fluency within one month), reconstructing path nodes accordingly to prioritize relevant modules and compress non-core content. This optimization mechanism increases the proportion of effective learning time to 89%, a 42-percentage-point improvement over traditional fixed-path models. In a 16-week comparative experiment, the experimental group using dynamic paths showed a significant 58.3% improvement in overall oral competence, much higher than the control group's 31.7%, while reducing learning frustration by 51%. The path optimization system also features predictive capabilities, using temporal models trained on historical data to anticipate learning bottlenecks (such as predicting comprehension obstacles in the "cross-cultural communication" module 3-5 weeks in advance) and automatically inserting preparatory knowledge linkages, achieving an interactive upgrade from "passive response" to "active intervention."

5. Technological Integration Innovation

5.1 Multimodal Interaction Enhancement

At the vocal modality level, voiceprint recognition technology is employed to precisely match learners' identities, ensuring personalized attribution of interaction data; in the textual modality, natural language generation (NLG) technology converts voice input into structured text summaries in real-time, facilitating learner review and analysis; visually, augmented reality (AR) technology superimposes layers of real-time translation subtitles, cultural background illustrations, and other information onto virtual dialogue scenarios—for instance, in simulating overseas shopping scenes, the AR interface automatically annotates professional vocabulary on product labels and provides gesture-interaction buttons. The action modality, through somatosensory capture devices, recognizes learners' body language, analyzes the rationality of their gestures during expression, and provides real-time corrective feedback. The core engine driving interaction enhancement lies in the integrated analysis of multimodal data. The system

employs cross-modal correlation algorithms to uncover potential links between different modalities: when voice recognition indicates a decline in pronunciation accuracy, it simultaneously analyzes the frequency of eye avoidance in the visual modality and physical stiffness in the action modality to determine whether anxiety is causing performance fluctuations, triggering emotional regulation mechanisms (e.g., switching to low-pressure scenarios). Experiments show that multimodal interaction improves learners' information reception efficiency by 37% and complex task completion rates by 51%. In advanced applications, multimodal technologies simulate non-verbal cues in cross-cultural communication (e.g., gestural semantic differences across cultures). Learners mimic target-culture body language via somatosensory devices, while the system evaluates its appropriateness in real-time. This "language-culture-action" synergistic training model expands the cultivation of intercultural communication skills from linguistic to multidimensional behavioral levels.

5.2 Human-Machine Collaboration Model

In foundational skill training, AI systems handle over 90% of repetitive tasks: voice recognition assesses pronunciation accuracy, dialogue systems generate standardized training questions, and virtual characters provide high-frequency scenario simulations, delivering more than five times the practice volume of traditional teaching within the same timeframe. For example, AI can evaluate the pronunciation of 20 learners within 30 minutes, a task requiring four hours for humans. At the strategic level, teachers design personalized "metacognitive training plans" based on AI-generated competence profiles—for instance, guiding learners to use self-questioning methods to enhance language monitoring abilities. Emotionally, teachers compensate for AI's interactive limitations through group discussions and role-playing, fostering authentic interpersonal dynamics; weekly two-hour human interactions increase learner motivation and retention by 28%. Culturally, teachers interpret the socio-cultural logic behind language, addressing tacit knowledge beyond AI's reach—one experimental class saw a 44% improvement in cross-cultural expression appropriateness after adding teacher-led cultural sessions (Yang, 2020). The depth of human-machine collaboration is reflected in dynamic role allocation: AI monitors learner progress curves to identify critical intervention points—e.g., triggering teacher alerts when a learner shows less than 5% improvement in the "argumentation" module over two weeks. Teachers then design targeted debate activities to break through bottlenecks. This "technical precision + human wisdom" model avoids AI's emotional gaps while resolving traditional teaching inefficiencies. Data shows that teams using this approach cover more advanced teaching goals (e.g., critical thinking) within the same class hours, accelerating oral competence growth by 29% compared to single-mode approaches, with 91% teaching satisfaction. The innovation of human-machine collaboration lies not in technological substitution but in functional reconstitution—"machines for repetitive training, humans for meaning construction"—establishing a new paradigm for smart education.

6. Conclusion

The interactive applications of artificial intelligence in English speaking instruction—encompassing voice recognition, dialogue systems, virtual characters, data-driven mechanisms, and technological integration—have established a novel interaction ecosystem characterized by real-time feedback, personalized guidance, and immersive experiences. These innovations transcend the constraints of traditional teaching, amplifying practice frequency, feedback efficiency, and contextual authenticity. Future development must prioritize enhancing technical precision and emotional intelligence, advancing human-machine collaboration, and steering oral language pedagogy toward intelligent, precise, and multidimensional interactive paradigms.

References

- Cai, B. (2024). The current status, development bottlenecks and future prospects of the application of artificial intelligence in English teaching at basic period from the perspective of "Internet+". *Journal of Artificial Intelligence Practice*, 7(3).
- Jiang, D., Pei, Y., Yang, G., et al. (2022). Research and analysis on the integration of artificial intelligence in college English teaching. *Mathematical Problems in Engineering*, 2022.
- Pitychoutis, M. K., & Rawahi, A. A. (2024). Smart teaching: The synergy of multiple intelligences and artificial intelligence in English as a foreign language instruction. *Forum for Linguistic Studies*, 6(6).

- Tong, Q., & Liu, S. (2025). Intelligent classroom environment: Application of internet of things and artificial intelligence in English teaching. *International Journal of High Speed Electronics and Systems*, (prepublish).
- Wang, J., Zhang, J., Chen, L., et al. (2024). Study of artificial intelligence-assisted English oral teaching. *Scientific Journal of Intelligent Systems Research*, 6(8), 1-8.
- Wei, X. (2024). Research on the application, challenges, and countermeasures of artificial intelligence in English teaching and learning. *Pacific International Journal*, 7(S1).
- Yang, G. (2020). The application of artificial intelligence in English teaching. *International Journal of Frontiers in Sociology*, 2(3).
- Zhang, Z. (2024). Practice exploration of artificial intelligence in higher vocational English teaching under informatization. *Evaluation of Educational Research*, 2(3).