DOI: 10.53469/jerp.2025.07(03).05

Construction and Practice of an AI-Based Personalized Training Mode for College English Listening and Speaking

Ting Li

Communication University of China, Nanjing, Jiangsu, China

Abstract: ***The rapid development of artificial intelligence technology has opened new avenues for the reform of college English listening and speaking teaching. In response to the shortcomings of traditional teaching models, such as severe homogeneity and insufficient attention to individual student differences, this paper constructs an AI-based framework for personalized English listening and speaking training. Relying on data-driven diagnostic analysis of learning situations, dynamic adaptation of learning paths, and the integration of multimodal resources, the framework builds a virtual simulation environment, integrating speech recognition and semantic analysis technologies to achieve high-frequency immersive language practice. Case studies indicate that AI technology significantly enhances students' oral fluency and listening comprehension, and stimulates learning initiative through an immediate feedback mechanism. However, current technology still faces challenges such as limited accuracy in speech recognition and insufficient algorithmic scenario adaptability. This paper proposes optimization strategies based on practical cases, providing theoretical reference for the digital transformation of education.***

Keywords: Artificial Intelligence, English Listening and Speaking Teaching, Personalized Training, Adaptive Learning**.**

1. Introduction

English listening and speaking instruction in higher education, as a crucial aspect of language skill development, has long been constrained by the homogenization dilemma of traditional teaching models, making it difficult to cater to students' diverse learning needs. Artificial intelligence technology provides an important breakthrough for addressing this issue. Through data-driven precise diagnosis, dynamic planning of adaptive learning paths, and the integrated application of multimodal resources, AI can infuse personalized and intelligent elements into English listening and speaking training. However, the process of technology empowering teaching still faces numerous practical challenges, hindering the deep integration of technology application and educational practice. This paper focuses on the intersection of artificial intelligence and college English teaching, systematically discussing the construction logic, implementation difficulties, and optimization paths of personalized training models for English listening and speaking. It aims to provide theoretical references and practical insights for breaking through traditional teaching bottlenecks and promoting the digital transformation of education.

2. The Significance of Personalized Training Models for English Listening and Speaking in Higher Education

2.1 Meeting the Practical Demand for Enhancing Students' Core English Competencies

2.1.1 The Strategic Value of English Listening and Speaking Skills in the Context of Globalization

The globalization process has accelerated the depth and breadth of cross-cultural exchanges, making English listening and speaking skills indispensable core competencies in international competition. According to the 2023 Global Talent Competitiveness Report, 85% of multinational corporations consider fluency in spoken English as a mandatory criterion for recruiting senior positions, while only 42% of college graduates in China meet the standard for oral proficiency, highlighting a disconnect between teaching and practice. Traditional teaching models rely on standardized textbooks and lack interactive training in real-life contexts, leading to prevalent issues such as "exam-oriented mute English" among students. AI technology, by constructing virtual simulation scenarios (such as international academic conferences, cross-border business negotiations, etc.), combined with speech recognition and natural language processing technologies, can simulate dynamic dialogues in multicultural backgrounds, providing students with high-frequency, immersive practical training. This technology-empowered training model not only compensates for the limitations of traditional classrooms but also directly responds to the demand for composite language talents in globalization.

2.1.2 Neglect of Individual Differences Among Students in Traditional Teaching Models

Differences in students' English listening and speaking abilities stem from the diversity of language foundations, cognitive styles, and learning motivations. Research shows that about 30% of students experience anxiety due to pronunciation issues, while 20% struggle to overcome listening comprehension barriers due to insufficient training intensity. Traditional "one-size-fits-all" teaching models fail to accurately identify individual shortcomings, such as issues with liaison weakening, intonation deviations, or logical expression confusion, which are often overlooked over time. AI-driven personalized systems generate student capability profiles through pre-diagnostic tests (such as speech feature analysis, grammar error clustering, etc.) and dynamically adjust training content based on reinforcement learning algorithms. For example, for students with weak pronunciation, the system can targetedly push “phoneme comparison training” and “stress and rhythm correction” modules; for those with listening comprehension difficulties, it enhances contextual reasoning abilities through multimodal input. This “tailored training” mechanism not only improves learning efficiency but also enhances students’ self-confidence through immediate feedback and positive incentives, effectively addressing the structural contradiction in traditional classrooms where “advanced students are underchallenged and struggling students cannot keep up [1].”

3. Practical Significance of Promoting Reform in College English Teaching

3.1 Transition from “Uniformity” to “Individualization”

College English teaching has long relied on standardized instruction models, neglecting cognitive differences and heterogeneous needs among students, which has affected the actual teaching effectiveness. AI technology, with its dynamic adaptation mechanism, can precisely identify students’ knowledge gaps and cognitive styles, establishing a new educational paradigm centered on learning. For example, reinforcement learning-based path planning models break down learning content into atomic ability units, utilizing nonlinear combinations to match individual development trajectories and realize “thousands of faces for thousands of people” training schemes. This transition not only grants students autonomy in learning but also shifts educational goals from standardized test scores to fostering students’ dynamic language abilities and cross-cultural thinking.

3.2 Technology Integration Enhances Teaching Efficiency and Quality

The deep integration of AI and education essentially involves using technology to reconstruct the underlying logic of the teaching process. In terms of efficiency, speech recognition and natural language processing technologies enable automated assessment and immediate feedback, freeing teachers from repetitive tasks and allowing them to focus on advanced instructional design (such as precise interventions based on heatmaps of group weaknesses). In terms of quality, the fusion of multimodal resources and adaptive recommendation algorithms optimizes resource allocation: resources such as “synchronized audio-visual analysis” are pushed to those with hearing impairments, while tasks like “semantic network construction” are designed for those with weak logical expression, significantly enhancing the targeting of training. Additionally, virtual simulation and augmented reality technologies break through the spatiotemporal constraints of traditional classrooms, constructing a ubiquitous language practice ecosystem. The AI-driven dynamic assessment system integrates process data (practice frequency, cognitive load fluctuations) and outcome indicators to build a multi-dimensional ability evaluation model, shifting teaching practice from experience-driven to scientific and refined evidence-driven [2].

4. Systematic Framework for the Personalized Training Mode of College English Listening and Speaking

Artificial intelligence, with its multidimensional data collection and analysis capabilities, has revolutionized the traditional teaching evaluation model that relies on experience. An AI-based English listening and speaking training system can capture dynamic learning data such as students’ voice characteristics, error clustering, and cognitive behavior trajectories in real-time, thereby constructing a personalized learning portfolio.

4.1 Multi-level Design of the Technical Architecture

4.1.1 Data Layer: The data layer serves as the foundation for system operation, responsible for multidimensional collection, storage, and management of learning data. The system integrates multimodal data sources such as voice input, text interaction, and behavior logs to build a dynamic learning situation database. This includes:

1) Voice Data: Collecting students’ pronunciation characteristics;

2) Behavior Data: Recording practice frequency, cognitive load fluctuations, and task completion paths;

3) Result Data: Storing test scores, clustering of error types, and ability enhancement trajectories. The distributed storage and privacy encryption technologies adopted by the data layer ensure data security and efficient retrieval.

4.1.2 Algorithm Layer: The algorithm layer serves as the intelligent hub of the system, encompassing three core functional modules:

1) Speech Recognition: Using deep neural networks (such as the Transformer model) to optimize the resolution accuracy of non-standard pronunciation, supporting recognition in complex scenarios such as liaison weakening and intonation deviations;

2) Path Planning: Inferring students’ knowledge states based on reinforcement learning and Bayesian networks, combined with the “Zone of Proximal Development” theory to delineate ability boundaries and dynamically generate nonlinear learning paths;

3) Feedback Generation: Utilizing Natural Language Processing (NLP) technology to analyze error patterns and generate voice spectrum comparison charts, immediate correction suggestions, or push micro-lecture resources for precise intervention.

4.1.3 Application Layer: The application layer provides interactive interfaces and functional services for teachers and students, including:

1) Virtual Simulation Scenarios: Simulating real-life contexts such as international academic conferences and business negotiations to support immersive language practice;

2) Personalized Dashboard: Visually displaying learning progress, ability maps, and feedback suggestions;

3) Multi-terminal Collaborative Platform: Supporting seamless switching between PCs, mobile devices, and VR equipment to construct a ubiquitous learning ecology.

4.2 Functional Realization of Core Modules

4.2.1 Pre-learning Diagnostic Module

Combining pre-voice tests and clustering analysis of grammatical errors to generate students' ability portraits. For example, voice characteristic analysis can identify weak phonemes in pronunciation, while grammatical error clustering can locate issues such as logical expression or tense misuse, providing data support for subsequent learning path planning.

4.2.2 Adaptive Learning Engine

Constructing a multi-objective optimization model based on reinforcement learning and collaborative filtering algorithms. The system dynamically adjusts training content through the Q-learning algorithm:

1) Priority pushing “audio-visual synchronous analysis” and “step-by-step dictation” tasks for those with hearing impairments;

2) Matching an “advanced scheme from phoneme correction to impromptu speech” for those with speaking difficulties;

3) The deep learning recommendation engine integrates group behavior data and achieves knowledge transfer and path optimization through matrix decomposition, breaking the rigid constraints of traditional teaching.

4.2.3 Intelligent Feedback System

Integrating speech recognition, sentiment analysis, and Natural Language Generation (NLG) technologies to enable immediate feedback and dynamic motivation. For example, when a pronunciation error is detected, the system generates a spectrum comparison chart and pushes correction videos; when anxiety is detected during listening training, it triggers calming practice suggestions. Meanwhile, gamification design (such as “ability badge” reward mechanisms) is introduced to enhance learning motivation and participation [3].

This framework constructs a personalized training ecology of “data-driven diagnosis - dynamic path planning - intelligent feedback loop” through the coordinated operation of technical levels and the linkage optimization of core modules, providing systematic support for enhancing students' language practical abilities and autonomous learning effectiveness.

5. Existing Issues in Constructing Personalized Training Models for College English Listening and Speaking

5.1 Challenges at the Technical Application Level

The current application of AI technology in college English listening and speaking training still faces significant technical bottlenecks, mainly manifested in the insufficient accuracy of speech recognition and natural language processing, as well as the limited adaptability of algorithms to complex linguistic scenarios. Speech recognition systems are prone to misjudgments when capturing non-standard pronunciations, continuous reductions, and intonation deviations, leading to distorted feedback and undermining the reliability of personalized training. Additionally, natural language processing models lack contextual sensitivity when dealing with multimodal interactions (such as emotional expressions and cultural metaphors), making it difficult to simulate the dynamics of real-life communication.

5.2 Realistic Contradictions in Teaching Implementation

The personalized training model for English listening and speaking driven by artificial intelligence faces multidimensional conflicts between technical logic and the traditional educational ecosystem during actual implementation, specifically manifested in the following three sets of contradictions: 1) The tension between teachers' subjectivity and technology's instrumentality. The existing teacher training system lacks course design for the deep integration of educational technology, resulting in a competency gap for teachers between technological application and teaching innovation. 2) The contradiction between students' cognitive habits and system compatibility. The passive learning mode shaped by traditional standardized training conflicts structurally with the autonomous exploration pathway advocated by AI. Some students, due to long-term dependence on unified instruction, struggle to adapt to the dynamic path planning of personalized systems, experiencing “choice anxiety” or “path loss,” which diminishes the expected effects of technology empowerment. 3) The blurred boundary between technological optimization logic and educational ethics.

5.3 Limitations of Resources and Evaluation Systems

The scarcity of high-quality corpora and multimodal training resources has become a critical factor constraining the development of personalized training models. Existing resources mostly focus on general scenarios, lacking specialized content for subdivisions such as academic discussions and cross-cultural communication, leading to a disconnection between training and actual needs. Furthermore, the evaluation system for personalized learning outcomes has not yet broken through the shackles of traditional quantitative indicators, overly relying on test scores or practice duration while neglecting implicit dimensions such as emotional state and cognitive strategies. The lack of multidimensional evaluation makes it difficult for the system to accurately identify students' deep learning needs, thereby affecting the pertinence and timeliness of dynamic feedback.

6. Practical Measures for Personalized Training Models in College English Listening and Speaking

6.1 Technical Path Optimization and Platform Construction

It is necessary to construct an intelligent technical framework integrating multimodal fusion to address issues such as insufficient speech recognition accuracy and weak algorithm generalization ability. For instance, introducing deep neural networks (such as the Transformer model) to optimize speech recognition systems, combined with transfer learning to enhance the system's ability to adapt to non-standard pronunciations; developing context-enhanced modules based on knowledge graphs, integrating semantic relation networks for cross-cultural scenarios, and strengthening natural language processing's ability to parse complex elements such as metaphors and emotions [4].

For example, the "AI English Listening and Speaking Intelligent Training Platform" developed in collaboration between Beijing Foreign Studies University and iFLYTEK uses the Transformer model to optimize the speech recognition engine, employs transfer learning technology to adapt to learners with different dialect backgrounds, avoiding pronunciation issues such as liaison and reduction. The system also incorporates a knowledge graph module, integrating semantic networks for scenarios such as academic conferences and business negotiations, supporting dynamic contextual reasoning. For instance, when students simulate an "international academic defense," the system accurately identifies logical flaws and pushes relevant case analysis videos. After the platform's launch, students' oral fluency improved by 23%, and listening error rates decreased by 18%, significantly outperforming traditional tools.

6.2 Integration of Teaching Resources and Scene Innovation

To address resource scarcity and scene monotony, it is necessary to collaborate with universities and industry institutionss to develop specialized corpora covering vertical scenarios such as academic discussions and business negotiations, including real dialogue recordings, audiovisual texts, and cultural background analysis databases. Utilizing generative AI (such as GPT-4) to automatically generate personalized training materials, combined with virtual reality (VR) to construct immersive interactive scenarios (such as simulations of international conferences), enhancing the authenticity of language practice.

For example, Shanghai Jiao Tong University, in collaboration with New Oriental Education & Technology Group, co-developed the "VR Cross-Cultural Communication Practical Training System," relying on an intelligent dialogue engine to construct interactive frameworks for scenarios such as business negotiations and academic discussions, integrating real corporate case resources to establish a 500-hour multimodal corpus. Students use virtual reality equipment to enter simulation scenarios such as "cross-border merger and acquisition negotiations," engaging in real-time negotiations with negotiation partners intelligently generated by the system, with speech recognition and emotion analysis technology dynamically assessing students' on-the-spot response capabilities. In a simulation of the "product pricing dispute" segment, students failed to capture metaphorical expressions in the opponent's cultural context, leading the negotiation to a stalemate. The system immediately matched relevant learning modules and pushed micro-courses on cross-cultural communication strategies. This project achieved an 89% coverage rate and was recognized as a "benchmark case of industry-education integration innovation" by China Education Daily.

6.3 Improvement of Dynamic Assessment and Feedback Mechanisms

To transcend the limitations of traditional quantitative evaluation, it is necessary to construct a three-dimensional dynamic assessment model encompassing "process, ability, and emotion." Integrate multimodal data (such as cognitive load fluctuations and speech emotion analysis) and eye-tracking technology to precisely capture implicit learning states. The feedback mechanism should integrate formative assessment and gamification design: set "ability badges" to reward periodic breakthroughs and use natural language generation (NLG) technology to provide personalized improvement suggestions. Simultaneously, establish an ethical review mechanism to regulate data collection boundaries, adopting federated learning technology to achieve collaborative data optimization under privacy protection, ensuring a balance between technology empowerment and educational ethics.

For example, Zhejiang University, in collaboration with Alibaba DAMO Academy, developed the "English Listening and Speaking Dynamic Assessment System," innovatively integrating eye-tracking and speech emotion recognition technology, capable of real-time capturing students' visual focus distribution and psychological stress changes during listening training. During "academic lecture listening" training, the system detected a student frequently staring at the vocabulary area and producing stiff pronunciations, immediately triggering the intelligent feedback mechanism: "It is recommended to infer vocabulary meanings based on context and try the 'relaxed following reading training' module." Test data showed a 31% reduction in users' tension levels and a 45% increase in the frequency of effective listening strategies, with this project selected as a typical case for the Ministry of Education's "AI+Education" innovation practice directory.

7. Conclusion

Artificial intelligence technology has provided a novel paradigm for the reform of English listening and speaking instruction in higher education. The AI-empowerment pathway proposed in this paper, relying on dynamic learning diagnostics, adaptive learning planning, and multi-modal resource integration, has significantly enhanced students' practical language skills and learning autonomy, serving as a practical example for the digital transformation of education. In the future, it is necessary to deepen interdisciplinary collaboration, optimize speech recognition accuracy and algorithm generalization capabilities, construct specialized corpora and multi-dimensional dynamic evaluation systems, and promote the deep integration of AI with language teaching. With the iteration of technology and the innovation of educational concepts, the AI-based personalized training model will gradually achieve a leap from “precise intervention” to “ecosystem construction,” infusing enduring momentum into the cultivation of compound language talents in a globalized context.

References

[1] Zhao Ying. Exploring the Digital Transformation of Pragmatic Competence Cultivation in College English Listening and Speaking Instruction [J]. The Road to Success, 2024, (24): 73-76.

[2] Xu Jia. Analysis of the Application of Multi-modal Teaching Mode in College English Listening and Speaking Courses [J]. Overseas English, 2022, (13): 110-111.

[3] Hong Jie. Inquiry into the Innovation of College English Teaching Based on Intercultural Communication Theory [M]. Xinhua Publishing House: 202107.174.

[4] Gao Xiuqin. On the Impact of Artificial Intelligence Education Products on College English Teaching [J]. Overseas English, 2021, (11): 111-112.