English Language Teaching and Linguistics Studies
ISSN 2640-9836 (Print) ISSN 2640-9844 (Online)
Vol. 8, No. 4, 2026
www.scholink.org/ojs/index.php/eltls

Original Paper

Exploration of Personalized English Learning Applications and Practices on Mobile Terminals Under Voice Interaction Technology

Jing Li[1]

[1] School of Translation Studies, Xi'an International Studies University, Xi'an 710128, China

Received: May 25, 2026 Accepted: July 26, 2026 Online Published: August 17, 2026

doi:10.22158/eltls.v8n4p123 URL: http://dx.doi.org/10.22158/eltls.v8n4p123

Abstract

The proliferation of mobile internet and the advancement of artificial intelligence have made voice interaction a significant breakthrough in the field of English language learning. This paper focuses on voice interaction technology and examines its application pathways and practical outcomes in personalized English learning on mobile terminals. Research indicates that the coordinated use of technologies such as automatic speech recognition and natural language processing can effectively support personalized learning needs across multiple dimensions, including vocabulary, grammar, and oral expression, while adaptive systems further enable dynamic matching of learning content and progress. On this basis, the paper constructs a multi-dimensional evaluation index system and proposes optimization recommendations addressing the technical limitations and application bottlenecks present in current practice.

Keywords

voice interaction technology, mobile learning, personalized English learning, adaptive system, oral training

Introduction

In recent years, the widespread adoption of smartphones has gradually made mobile learning one of the primary contexts for language acquisition, and the maturation of voice interaction technology has injected new vitality into this setting. Unlike traditional text-and-image interaction, voice interaction can engage more directly with English listening and speaking training, breaking the constraints of time and space and enabling learners to complete meaningful language input and output during fragmented moments of the day. However, existing mobile English learning applications still fall noticeably short in terms of personalization. Most products remain at the level of content delivery and struggle to respond in real time to individual learners' pronunciation characteristics, proficiency levels, and learning habits. Against this backdrop, this paper focuses on how voice interaction technology can genuinely drive the realization of personalized learning, conducting a systematic investigation across three dimensions: the mechanism of technology integration, the construction of practical models, and the optimization of effect evaluation, with the aim of providing reference for innovative practices in mobile English learning.

1. The Integration Mechanism of Voice Interaction Technology and Mobile English Learning

1.1 Core Functions of Voice Interaction Technology in English Learning

The value of voice interaction technology in English learning is primarily reflected across three dimensions: immediate feedback, two-way dialogue, and situational simulation. Traditional English learning has long been dominated by one-way input, leaving learners without an effective channel to verify their language output. The introduction of voice interaction technology breaks this limitation. Through automatic speech recognition (ASR), learners' pronunciation is captured and analyzed in real time, with scoring and error correction feedback returned within milliseconds, significantly improving the efficiency of oral practice. At the same time, virtual language partner functions built on dialogue management modules can simulate authentic communicative scenarios, guiding learners through task-based language interaction and fundamentally addressing the longstanding problem of learning without use (Chen, An, Meng et al., 2025). The core contribution of voice interaction technology to English learning lies in transforming passive reception into active output, providing an operationally accessible entry point for personalized learning.

1.2 The Technical Support Framework for Personalized Mobile Learning

The realization of personalized English learning on mobile terminals depends not on any single technological breakthrough, but on the coordinated operation of multiple technical layers working in concert. At the hardware level, the high-sensitivity microphones and processing chips integrated into modern smartphones provide the foundational conditions for local speech capture and preliminary processing, reducing excessive reliance on network connectivity and improving the learning experience in unstable network environments. At the data level, every interaction a learner has with the application generates analyzable traces, including speech samples, response sequences, time-on-task, and error type distributions. Once structured and stored, these data form a dynamically updated personal learning profile that serves as the core basis for personalized recommendations. At the algorithmic level, models such as Bayesian Knowledge Tracing and Deep Knowledge Tracing continuously model the probability of a learner's mastery across individual knowledge points, and in combination with forgetting curve theory, intelligently schedule review content to avoid inefficient repetition and knowledge gaps. Recommendation algorithms further draw on learner profiles to select materials from content libraries that best match current needs in terms of difficulty, topic, and format. These technical components reinforce one another in a layered structure, forming the operational backbone of personalized mobile learning and translating the educational ideal of teaching to individual needs into executable system logic.

1.3 The Collaborative Role of Speech Recognition and Natural Language Processing in English Teaching

Speech recognition and natural language processing are the two core technologies underpinning intelligent English instruction on mobile terminals. Each has its distinct focus, yet the two are highly complementary, and their coordination is essential to achieving a deeper understanding of learners' language performance. Speech recognition is responsible for converting acoustic signals into analyzable text, a process involving acoustic modeling, language model constraints, and decoding strategies, where adaptability to accent, speech rate, and background noise directly determines the reliability of the output. Building on this, natural language processing conducts deep linguistic analysis of the transcribed text, covering dimensions such as appropriateness of vocabulary choice, grammatical accuracy, and semantic coherence, extending the system's evaluative capacity from the phonetic level to the level of language use. In oral assessment scenarios, this collaboration enables the system to deliver comprehensive feedback that accounts for both speech quality and linguistic quality, rather than producing a single score based solely on phonetic similarity. In grammar assistance scenarios, NLP modules can identify sentence-structure transfer errors common among Chinese-speaking learners and provide contextually sensitive revision suggestions, helping learners understand the linguistic reasoning behind their mistakes rather than simply marking responses as right or wrong (Wang, 2023). The deep integration of these two technologies gives the system a preliminary capacity to understand learners' communicative intent, laying a solid linguistic-analytical foundation for subsequent personalized content delivery, learning pathway planning, and dynamic difficulty adjustment.

2. Practical Models for Personalized English Learning on Mobile Terminals

2.1 Personalized Vocabulary and Grammar Learning Pathways Based on Voice Interaction

Vocabulary and grammar form the two foundational pillars of English proficiency, and the introduction of voice interaction technology has made the personalized learning pathways for both dimensions considerably more concrete and operationally feasible. In vocabulary learning, traditional methods compress the memorization task into the repeated visual presentation of written symbols, long neglecting the phonetic dimension and resulting in a situation where many learners can recognize words but cannot produce them orally. With voice interaction integrated, the system requires learners to read target vocabulary aloud within context, using read-along comparison to instantly verify pronunciation accuracy, while contextual example sentences help establish simultaneous memory links across the auditory, orthographic, and semantic dimensions, meaningfully strengthening retention. The system also dynamically adjusts the frequency and format of vocabulary review based on each learner's error history and forgetting patterns, applying targeted reinforcement to persistently problematic words while reducing repetition density for already-mastered items, concentrating learning resources precisely where they are needed. In grammar learning, voice interaction brings a substantive change to practice formats. Rather than completing multiple-choice questions or correction exercises alone, learners practice applying grammatical rules through oral sentence construction, with the system using syntactic analysis to evaluate the structural accuracy of spoken output and providing explanations tied to specific grammar points, steering grammar learning away from rote memorization and toward actual use. Taken together, voice interaction enriches vocabulary and grammar learning with broader sensory engagement and more timely feedback, while providing a richer data foundation for ongoing personalized adjustment.

2.2 Personalized Strategies for Oral Training and Pronunciation Error Correction

Oral training is the application scenario in which voice interaction technology holds the clearest advantage, and personalized pronunciation error correction is both the most technically demanding and the most distinctively valuable component within that training system. Effective pronunciation correction does not simply inform learners that their pronunciation is inaccurate; it must locate the specific type and position of the error and provide actionable guidance for improvement (Liu, 2025). The system first conducts multi-layered comparison of the learner's phonemic production, tonal contour, connected speech patterns, and pause rhythm against a standard pronunciation database, using deviation calculations relative to the reference model to precisely identify where pronunciation problems lie and generate granular error reports. Given that learners from different native language backgrounds exhibit clearly patterned differences in pronunciation errors, the system can apply differentiated error correction models tailored to specific native language groups. For Chinese native speakers, for instance, the system prioritizes frequent transfer errors such as aspiration contrasts, vowel length distinctions, and sentence-final intonation patterns, substantially improving the precision of feedback. At the level of training design, the system generates personalized practice sequences based on each learner's correction history and error frequency, applying progressive targeted intervention for persistent errors through a cycle of listening discrimination, imitation, contrastive analysis, and re-production, gradually building accurate phonetic muscle memory. At the same time, training content is not confined to isolated syllables or words but is embedded within authentic conversational contexts, allowing learners to improve their pronunciation naturally in the course of completing communicative tasks and avoiding the erosion of motivation that comes with decontextualized mechanical drilling.

2.3 Application of Adaptive Learning Systems in Developing English Listening and Speaking Proficiency

Adaptive learning systems represent the most highly integrated and systemically valuable technical form within personalized mobile English learning practice. Their defining feature is the use of real-time data as a driver to continuously adjust the content, difficulty, and pacing of listening and speaking training so that the learning process remains aligned with the learner's actual ability boundary. In listening training, the system analyzes a learner's comprehension accuracy across different materials, response latency, and the linguistic characteristics of error clusters, then dynamically matches speech rate, accent type, lexical density, and topic domain to ensure that input materials consistently fall within the learner's current zone of comprehensible input. Materials that are too difficult lead to repeated comprehension failures and accumulated frustration, while materials that are too easy produce little meaningful development; the value of the adaptive mechanism lies precisely in its capacity to continuously locate the optimal balance between these two extremes. In oral proficiency development, the adaptive system arranges dialogue tasks in levels according to communicative complexity, ranging from simple information-confirmation exchanges to mid-level tasks requiring description, explanation, and opinion expression, through to higher-order tasks demanding argumentation and persuasion. The system dynamically adjusts task level based on the quality of the learner's actual performance and triggers vocabulary hints or sentence-structure scaffolding at critical communicative junctures to help learners overcome expressive barriers (Huang & Li, 2025). As learning data accumulates over time, the system's portrait of each learner's abilities becomes increasingly three-dimensional, enabling it to anticipate weak areas and intervene proactively rather than merely responding to errors after they appear, shifting the cultivation of listening and speaking proficiency away from linear curriculum progression and toward a flexible, individually tailored developmental pathway.

3. Effect Evaluation and Optimization of Personalized English Learning Driven by Voice Interaction

3.1 A Multi-Dimensional Evaluation Index System for Personalized Learning Outcomes

Evaluating the effectiveness of personalized English learning driven by voice interaction technology requires moving beyond single-metric score measurement and establishing a comprehensive index system capable of reflecting the full picture of the learning process. The linguistic competence dimension serves as the core layer of evaluation, with specific indicators including the magnitude of vocabulary growth across learning stages, changes in grammatical accuracy, improvements in oral fluency, and the degree of progress in pronunciation standard. These indicators can be obtained through stage-based assessments embedded in the system and longitudinal comparison of daily practice data, and they carry strong quantitative operability. The learning behavior dimension focuses on the quality of the learning process itself, encompassing the consistency of learning frequency, task completion rates, the proportion of self-initiated review, and the recurrence rate of the same types of errors. These process-oriented indicators reveal the depth of learner engagement and self-directed learning capacity, making them important reference points for predicting long-term outcomes (Ma, 2025). The learning experience dimension turns attention to learners' subjective perceptions, covering their acceptance of the accuracy of voice interaction feedback, their sustained willingness to use the system, and the sense of self-efficacy they develop through learning. Data from this dimension often determines whether learners are willing to persist with the system over time. The three dimensions are mutually independent yet interconnected: the linguistic competence dimension reflects outcomes, the learning behavior dimension reflects process, and the learning experience dimension reflects motivation. Only by bringing all three together can a complete assessment of personalized learning effectiveness be formed, and only then can a sufficiently comprehensive and targeted basis be provided for subsequent optimization of learning plans.

3.2 The Dynamic Adjustment Mechanism of Learner Feedback Data on Personalized Plans

The true value of a personalized learning plan lies not in how precisely it is configured at the outset, but in whether it can continuously sense change and respond accordingly throughout the learning process. Realizing this capacity depends on a well-functioning mechanism for collecting feedback data and making dynamic adjustments. System-level adjustment is driven by behavioral data, with backend algorithms continuously tracking learners' performance trajectories across individual knowledge points (Chen, 2025). When the error rate on a particular type of task remains consistently high across multiple consecutive practice sessions, the system automatically determines that the content exceeds the learner's current capacity and immediately triggers a process of difficulty reduction or targeted foundational reinforcement. Conversely, when indicators in a given competence dimension show a stable upward trend, the system accordingly raises the challenge gradient in that area to maintain ongoing activation of learning motivation. Relying solely on behavioral data for adjustment has inherent limitations, however, because behavioral data reflects what learners do rather than how they feel: it cannot directly convey emotional states or levels of learning satisfaction. To compensate for this gap, the system collects learners' subjective feedback at key learning junctures through brief voice questionnaires or rating scales, incorporating data on emotional state, fatigue level, and content preferences into its adjustment parameters, preventing the system from overlooking learners' genuine experience through overreliance on objective metrics. The coordinated operation of objective behavioral data and subjective feedback data ensures that adjustments to personalized plans are both evidence-based and attentive to human needs, forming a complete iterative cycle of execution, collection, analysis, and adjustment that progressively brings each learner's plan closer to its optimal form.

3.3 Critical Reflections on Current Practice and Future Development Pathways

The application potential that voice interaction technology has demonstrated in personalized mobile English learning is genuine, but treating it with uncritical optimism and sidestepping real problems serves the field's development poorly. At the technical level, current speech recognition systems still exhibit noticeable fluctuations in accuracy when handling non-standard accents, high-noise environments, and the rapid, continuous flow of natural speech. When a system misidentifies a learner's perfectly normal utterance as an error and delivers inappropriate feedback, the consequences extend beyond degraded training outcomes to include erosion of the learner's trust in the system, which represents one of the core bottlenecks constraining the voice-interactive teaching experience. At the data level, building high-quality personalized models depends on large-scale annotated data, yet sufficiently detailed annotations capturing the vocabulary distribution, pronunciation error patterns, and learning habits of specific learner subgroups remain inadequate, limiting the generalization capacity of personalized algorithms when they encounter learners from less-represented groups and reducing the precision of recommendations. At the application level, a number of mobile English products suffer from a clear disconnect between their technical features and their pedagogical objectives. Voice functionality is layered into applications as a selling point rather than being organically integrated with specific learning goals, and learners end up expending significant cognitive effort navigating complex voice interaction processes while gaining limited actual language benefit. Looking to the future, as large language models continue to advance in their capacity for language understanding and generation, voice interaction systems are poised to deliver increasingly natural and fluid conversational experiences, moving beyond simple question-and-answer drills toward what might genuinely be described as a language exchange partner. The incorporation of multimodal interaction will further enrich the authenticity of learning contexts, while advances in edge computing hold the promise of enabling more precise real-time speech analysis with reduced dependence on network connectivity. These technological trajectories are not remote possibilities, but translating them into effective personalized English learning tools will require much deeper collaboration between technology developers and language education researchers, building a closer and more sustained dialogue between engineering implementation and pedagogical design.

4. Conclusion

This paper has conducted a systematic examination of the application and practice of personalized English learning on mobile terminals driven by voice interaction technology, organized across three dimensions: integration mechanisms, practical models, and effect evaluation. The findings indicate that the coordinated use of speech recognition and natural language processing provides a solid technical foundation for personalized English learning, while the integration of adaptive systems enables learning pathways across vocabulary, grammar, listening, and speaking to be dynamically adjusted in response to learners' actual performance, fundamentally addressing the longstanding limitations of fixed content and delayed feedback in traditional mobile English learning. At the level of effect evaluation, a multi-dimensional index system covering linguistic competence, learning behavior, and learning experience, combined with a dual-channel feedback data mechanism for dynamic adjustment, provides a relatively complete operational framework for the continuous optimization of personalized plans. That said, insufficient speech recognition accuracy, limited annotated data resources, and low coupling between technical features and pedagogical goals remain practical obstacles constraining the quality of current implementations and should not be obscured by an atmosphere of uncritical technological optimism. Looking further ahead, advances in large language models, multimodal interaction, and edge computing will open broader possibilities for personalized mobile English learning, but whether those possibilities can be converted into genuine learning gains will ultimately depend on whether technological capability and pedagogical wisdom can achieve true and deep integration in practice.

References

Chen, F. (2025). Analysis of virtual digital human intelligent voice interaction control technology. *Information and Computer*, *2025*(7), 29-31.

Chen, S. T., An, K., Meng, F. L. et al. (2025). System design of children's companion robot based on mobile IoT technology. *Internet of Things Technologies*, *15*(6), 97-101, 106.

Huang, J. F., & Li, Q. (2025). Research on application of AI technology in practical English daily communication training teaching. *Computer Procurement*, *2025*(34), 278-280.

Liu, H. (2025). Research on application of artificial intelligence technology in optimizing classroom interaction in junior high school English. *Future Scientist*, *2025*(34), 23-24.

Ma, X. M. (2025). Research on application of AI personalized learning in traditional culture teaching in primary school English. *China Selected Articles (Traditional Culture Teaching and Research)*, *2025*(13), 67-69.

Wang, M. S. (2023). Research on application strategies of English nursery rhymes in phonetic teaching under AI empowerment. *Mangrove*, *2023*(24), 0079-0081.