European Journal of Pedagogical Initiatives and Educational Practices
ISSN (E): 2938-3625
Volume 4, Issue 2, February - 2026

USING ARTIFICIAL INTELLIGENCE TO IDENTIFY SPEAKING ERRORS IN REAL-TIME

Laylo Matyakubova
Trainee- Teacher,
Uzbekistan State World Languages University

Abstract
This article explores how artificial intelligence (AI) enables real-time detection and correction of speaking errors by using advanced speech recognition, machine learning, and natural language processing technologies. It examines the principles behind automatic speech recognition (ASR), pronunciation error identification, real-time feedback mechanisms, and educational applications. By analyzing current research and practical systems, the article highlights both the opportunities and challenges of AI-driven speaking error identification, especially in language learning and communication accuracy enhancement.

Keywords: AI speech recognition, real-time error detection, pronunciation analysis, automatic feedback, natural language processing, machine learning, speaking proficiency, language learning technology.

Introduction

Artificial intelligence has fundamentally transformed the way spoken language can be analyzed, evaluated, and improved. For decades, speech assessment depended almost entirely on human evaluators who listened, interpreted, and judged spoken performance. While human assessment remains valuable, it is inherently limited by subjectivity, time constraints, and scalability challenges. The emergence of advanced automatic speech recognition systems has changed this paradigm, enabling machines not only to transcribe speech but also to interpret linguistic patterns and detect deviations in real time. The ability of AI to identify speaking errors as they occur represents one of the most promising intersections of machine learning, linguistics, and education.

In addition, artificial intelligence is profoundly transforming the ways we understand, analyze, and improve human speech. At the core of many modern applications is automatic speech recognition (ASR) — systems that convert spoken language into text and analyze phonetic, syntactic, and prosodic features of speech. The evolution of ASR, underpinned by deep learning models and neural network architectures, has made real-time detection of speaking errors feasible with increasing accuracy and responsiveness.

Real-time speaking error detection refers to the capability of an AI system to listen to speech as it is being produced, identify deviations from expected pronunciation, grammar, or pattern norms, and provide instantaneous feedback. Historically, assessment of spoken performance was limited to human raters or delayed review sessions, both of which are slow, subjective, and resource-intensive. In contrast, AI-powered systems offer objective, immediate, and scalable

alternatives, unlocking new potential for applications such as second language learning, public speaking coaching, speech therapy, and assistive communication tools.

The technological foundation of real-time error detection is advanced speech recognition models that use deep architectures like transformers, recurrent neural networks, and encoder-decoder frameworks. These models are trained on large annotated datasets containing diverse speech samples with accents, noise conditions, and dialectal variations. Modern ASR systems such as OpenAI's Whisper model demonstrate high transcription accuracy across multiple languages and noisy environments, helping to ensure that real-time detection systems correctly capture spoken input before performing error analysis.

Once speech is transcribed, AI models compare the recognized text and phonetic output with reference pronunciations and linguistic rules. Algorithms can detect pronunciation errors, such as incorrect articulation of phonemes, stress misplacement, or prosodic irregularities. Research shows that integrating acoustic, phonetic, and linguistic embeddings improves the identification of subtle mispronunciations and boosts diagnostic accuracy.

Real-time speech analysis systems typically use advanced machine learning algorithms and neural network models to classify identified mistakes into specific categories, such as sound substitutions, omitted elements, added segments, or broader phonetic inconsistencies. In educational contexts, these systems do more than simply point out errors. They record how often certain mistakes occur, analyze patterns in learner performance, and produce structured, meaningful feedback. Based on this analysis, AI can suggest targeted practice activities and adjust the level of difficulty, creating individualized learning trajectories that respond to each student's abilities, progress rate, and particular areas requiring improvement.

The application of real-time AI detection extends beyond educational environments. In clinical domains, AI also detects speech disorders and disfluencies. For example, models like StutterNet use time-delay neural networks to identify patterns associated with stuttering without requiring external linguistic models. Such systems can support speech therapists by automatically flagging disfluencies for further human evaluation. Real-time feedback is crucial for learner engagement and error correction efficacy. Feedback may be delivered audibly, visually, or via mobile applications that highlight phoneme-level discrepancies between learner speech and ideal references. An emerging trend is the development of interactive AI tutors that act like personal coaches, combining error detection with suggestions, examples of correct pronunciation, and practice drills.

Despite significant advances, several challenges remain. Latency and computational load can limit the responsiveness of real-time systems, especially on resource-limited devices. Ensuring accuracy under diverse accents, background noise, and spontaneous speech is still a research frontier. Additionally, real-time systems must be designed with careful attention to ethical considerations, including privacy of user speech data and transparency about how feedback is generated.

Another important dimension of real-time AI-based speaking error detection is the role of adaptive learning analytics and longitudinal performance tracking. Modern AI systems are increasingly designed not only to identify isolated pronunciation or grammatical errors, but also to monitor patterns over time. By collecting structured speech performance data across multiple sessions, AI can detect recurring error types, measure improvement rates, and predict areas of

persistent difficulty. This predictive capability allows systems to shift from reactive correction to proactive guidance, suggesting targeted exercises before errors become fossilized habits.

Furthermore, recent developments in reinforcement learning have enabled feedback systems to optimize how and when corrections are delivered. Research suggests that excessive correction may overwhelm learners, while insufficient feedback slows progress. AI models can dynamically adjust feedback frequency based on learner responsiveness, engagement levels, and cognitive load indicators. Some systems also integrate user-specific variables such as native language background, age, and proficiency level to personalize correction strategies.

A growing area of research focuses on cross-linguistic phonological analysis in AI-driven speech systems. By examining the sound structure of a learner's first language and comparing it with the phonetic patterns of the target language, artificial intelligence can predict which sounds are likely to cause difficulty. This allows the system to design preventive pronunciation exercises and tailored practice tasks before consistent errors develop. Such advancements show that real-time speaking error detection is moving beyond basic correction and becoming a comprehensive, data-informed framework for continuous speech improvement and personalized language development support.

The future of real-time AI-driven speaking error detection includes integration with augmented and virtual reality environments, enabling immersive language practice scenarios, and further refinement of contextual understanding through natural language processing. With ongoing research and improving computational capabilities, AI systems are poised to become indispensable tools for enhancing speaking proficiency, reducing communication barriers, and improving access to language learning resources worldwide.

Taking everything into consideration, the use of artificial intelligence to identify speaking errors in real time represents a transformative development in language education, speech therapy, and professional communication. Through advanced speech recognition, deep learning, and natural language processing, AI systems can provide immediate, objective, and personalized feedback. Although technical, ethical, and pedagogical challenges remain, ongoing research and innovation suggest that AI-driven speaking assessment will become increasingly sophisticated and widely adopted. The collaboration between human expertise and intelligent systems offers unprecedented opportunities to improve spoken communication skills across diverse contexts.

References

1. Bhattad, P., & Jain, V. (2025). Artificial intelligence to detect voice disorders: An AI-supported systematic review of accuracy outcomes. Journal of Voice.
2. Nabijeva, S. R. (2026). The role of AI algorithms in detecting students' pronunciation errors. International Journal of Artificial Intelligence.
3. OpenAI. (2022). Whisper (speech recognition system). Wikipedia.
4. Ye, W., Mao, S., Soong, F., Wu, W., Xia, Y., Tien, J., & Wu, Z. (2021). An approach to mispronunciation detection and diagnosis with acoustic, phonetic and linguistic (APL) embeddings.
5. Author Name. (2025). Interactive AI assistant for oral error analysis and feedback. HSE University Thesis.

6. Nabiyeva, D., & Abduramanova, D. V. (2025). Leveraging AI to analyze ESL learners' speech patterns across proficiency levels. The Lingua Spectrum.
7. Sheikh, S. A., Sahidullah, M., Hirsch, F., & Ouni, S. (2021). StutterNet: Stuttering detection using time delay neural network.