



LEVERAGING AI TO ANALYZE ESL LEARNERS' SPEECH PATTERNS ACROSS PROFICIENCY LEVELS

Diyora NABIYEVA

Student, Chirchik State Pedagogical University

di.yora22072002@gmail.com

Scientific adviser: **Diana Valeryevna ABDURAMANOVA**

Head of the English theory and practice department,

Doctor of Philosophy in Philological Sciences (PhD), act.assoc.prof.

Chirchik State Pedagogical University

d.abduramanova@cspi.uz

Abstract. *This study investigates the use of artificial intelligence (AI) to analyze speech patterns in English as a Second Language (ESL) learners across different proficiency levels, focusing on pronunciation, vocabulary, and syntactic errors. Traditional methods of evaluating learner speech, such as manual transcription and teacher feedback, are often time-consuming, subjective, and difficult to scale. By leveraging automatic speech recognition (ASR) and natural language processing (NLP) tools, this research aims to detect and categorize common speech errors, comparing their frequency and types among beginner, intermediate, and advanced learners. Hypothetical findings suggest that beginners exhibit the highest error rates, particularly in segmental pronunciation, vocabulary misuse, and syntactic constructions, while intermediates show improvement with occasional errors in complex structures, and advanced learners display relatively low error rates, with subtle mistakes in nuanced vocabulary and advanced syntax. AI detection is expected to align well with human annotations for clear pronunciation and grammatical errors, though subtler aspects such as intonation, rhythm, and pragmatics may show lower agreement. The study highlights the potential of AI-assisted analysis to provide objective, scalable feedback, inform curriculum design, and enhance personalized instruction. Limitations of AI, such as sensitivity to nonnative accents and subtle pragmatic errors, are acknowledged, suggesting that AI should complement, rather than replace, human evaluation in ESL learning.*

Keywords: *Artificial intelligence, ESL learners, speech analysis, pronunciation errors, vocabulary errors, syntactic errors, automatic speech recognition, natural language processing, language proficiency, AI-assisted language learning*

INTRODUCTION

Automatic Speech Recognition (ASR) – a branch of artificial intelligence (AI) – offers a promising way to support second language (L2) learning. In the context of English as a Second Language (ESL), learners often struggle with pronunciation, vocabulary use, and syntax, especially at lower proficiency levels. Traditionally, language teachers rely on manual feedback and error correction, which may be subjective and time-consuming, particularly in large classes. AI-based tools can help overcome these limitations by providing scalable, objective feedback on spoken production. For instance, a recent



meta-analysis of ASR-assisted pronunciation training reported a medium overall effect ($g = 0.69$), suggesting that ASR can significantly improve ESL/EFL learners' pronunciation when used appropriately.

Nevertheless, most research on ASR in language learning focuses narrowly on pronunciation, leaving other aspects – such as vocabulary misuse or grammatical errors in spontaneous speech – relatively underexplored. Furthermore, little attention has been paid to how error patterns vary across learners of different proficiency levels. Understanding these patterns could guide more targeted teaching and personalized feedback, potentially improving learning outcomes and learner autonomy.

This study aims to fill this gap by exploring whether AI can reliably detect pronunciation, vocabulary, and syntactic errors in ESL learners' spoken English, and whether the frequency and types of errors differ systematically across learners of varying proficiency (beginner, intermediate, advanced). The core research questions are: Can ASR and associated natural language processing (NLP) tools detect different types of errors in learner speech? What error patterns emerge at different proficiency levels? And how can these insights inform ESL teaching and assessment?

METHODS

Participants will be drawn from an ESL learner pool, grouped into three proficiency levels aligned roughly with common frameworks (e.g., beginner, intermediate, advanced), and determined via a reliable placement or proficiency test. Each participant will complete a set of speaking tasks under controlled conditions. The tasks will include a short reading passage (to control for content), a free-speech or picture-description task (to elicit spontaneous language), and a prompted dialogue or Q&A (to simulate conversational use). All speech will be recorded using good-quality audio equipment in quiet conditions.

The recorded speech samples will then be processed through an ASR system to generate text transcripts. For pronunciation analysis, forced alignment or phoneme-level analysis may be used, where possible, an AI-based pronunciation evaluation model or phoneme comparison algorithm will detect segmental and suprasegmental errors. For vocabulary and syntax analysis, the ASR output will be fed into NLP tools (or custom scripts) to detect lexical misuse (wrong word choice or word form), incorrect word usage, and grammatical/syntactic errors (e.g., word order, agreement, tense). An error-categorization framework will classify errors into categories like pronunciation (segmental, suprasegmental), lexical, syntactic, and possibly fluency or pragmatic.

For reliability, a subset of recordings will be annotated manually by human raters, who will mark errors independently. These human annotations will serve as a "gold standard" against which the AI-detected errors will be compared. Agreement metrics (e.g., accuracy, precision/recall for error detection) will evaluate how well the AI performs relative to human judgment. Quantitative analyses will compute error rates (e.g., number of errors per 100 words or per utterance) in each category for each participant and aggregate by proficiency level. Statistical comparisons (e.g., ANOVA) will assess whether differences in error rates across proficiency groups are significant. Qualitative analysis may look at representative examples of errors typical for each proficiency level, including common mispronunciations, vocabulary misuse patterns, and syntactic mistakes.

RESULTS



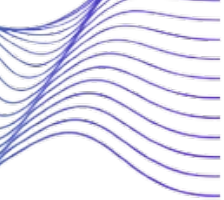
Section-1: Artificial Intelligence in Language Teaching

It is anticipated that lower proficiency learners, typically classified as beginners, will demonstrate comparatively high error rates across all analyzed categories of spoken language. Pronunciation errors are expected to be the most frequent among this group, particularly segmental errors involving misarticulation of individual phonemes or substitution of native language sounds for English sounds. These learners are also likely to produce limited or incorrect vocabulary, often relying on simple or memorized words and phrases, and may misuse basic lexical items due to insufficient familiarity with appropriate word forms or collocations. Syntactic errors are similarly expected to be prevalent, with common mistakes including incorrect word order, improper verb tense usage, and difficulties with subject-verb agreement. Intermediate learners, in contrast, are anticipated to show fewer pronunciation errors as their articulatory control improves and exposure to authentic language increases. Their vocabulary usage is expected to expand, though occasional lexical errors or word-choice problems may still occur, particularly with less common or contextually nuanced words. Syntactic mistakes are also expected to persist but tend to be more moderate, often arising in complex sentences or structures that require greater grammatical control. Advanced learners are likely to exhibit low overall error rates in pronunciation and fundamental syntactic structures. However, errors may still appear in more complex or less frequently used constructions, and vocabulary mistakes may be subtle, such as miscollocations or nuanced word misuse. When comparing AI-detected errors with human annotations, good agreement is expected for clear, observable mistakes, such as segmental pronunciation errors, overt lexical mistakes, or standard syntactic violations. More subtle features, including intonation, rhythm, pragmatics, or acceptable but nonnative constructions, may yield lower AI-human agreement. Overall, a downward trend in error frequency across proficiency levels is expected, with beginners displaying the most errors, intermediates fewer, and advanced learners the least, highlighting the developmental trajectory of spoken language proficiency in ESL learners.

DISCUSSION

If results align with expectations, this study will demonstrate the viability of using AI tools to analyze learner speech beyond pronunciation alone, extending into vocabulary and syntax, and capturing patterns across proficiency levels. Such findings could have important pedagogical implications. Teachers and curriculum designers could use AI-based analysis to identify common learner difficulties per proficiency stage and tailor instruction accordingly – for example, focusing on pronunciation and basic vocabulary with beginners, and shifting to complex syntax or nuanced vocabulary with advanced learners. AI-assisted feedback systems could complement traditional teaching, offer personalized feedback at scale, and promote learner autonomy, especially where teacher time is limited.

However, some limitations must be acknowledged. ASR systems are typically trained on native speaker speech or standard accents, their performance on non-native or accented speech may suffer. Indeed, recent research has shown that ASR systems can demonstrate higher word error rates with L2 English or accented speakers, especially when phonetic variations are underrepresented in training data. This raises questions about the reliability of AI detection for non-native speech, and suggests that manual



correction or at least validation remains necessary. Furthermore, AI may struggle with suprasegmental features (intonation, rhythm, stress) or pragmatic appropriateness – areas essential for communicative competence. Also, overreliance on AI feedback could reduce human interaction and neglect the social aspects of language learning.

Therefore, while AI-based analysis can offer valuable insights and scalability, it should be regarded as a complement rather than replacement for human judgment. Future work could explore improving ASR and NLP models for non-native speech, integrate prosodic and pragmatic evaluation, and test the effectiveness of AI-informed feedback in real ESL classrooms over longer periods.

CONCLUSION

This proposed study outlines how AI, in the form of ASR and NLP tools, can be leveraged to analyze ESL learners' spoken English at different proficiency levels, revealing patterns of pronunciation, vocabulary, and syntax errors. By grouping learners into proficiency levels and comparing error rates, the research seeks to provide empirical evidence of how learner difficulties evolve. The potential pedagogical benefits – including personalized feedback, scalable assessment, and informed curriculum design – make this approach a promising addition to ESL teaching. Yet limitations in current AI technology, especially regarding non-native accents and nuanced aspects of language, call for cautious application and further development. Overall, the study supports the vision of integrating AI as a powerful research and teaching assistant, paving the way for hybrid learning environments where human and machine feedback complement each other.

REFERENCES

1. Febrianti, Rizky Liliani. "The Use of Pronunciation Training Based on Automatic Speech Recognition of ELSA Speak Application for EFL Students." *RETAIN: Journal of Research in English Language Teaching*, vol. unknown, 2024, pp. – .
2. Farrús, Mireia. "Automatic Speech Recognition in L2 Learning: A Review Based on PRISMA Methodology." *Languages*, vol. 8, no. 4, 2023, p. 242, doi:10.3390/languages8040242.
3. Babekir, Ali Hussein Salim. "Speech Recognition and Pronunciation Apps and EFL Learners: A Study of Efficacy in Developing Speaking Skills." *Migration Letters*, vol. 20, no. S2, 2023, pp. 883–893, doi:10.59670/ml.v20iS2.3741.
4. McGuire, Michael. "Automatic Speech Recognition for Non-Native English: Accuracy and Disfluency Handling." 2025, arXiv, arXiv:2503.06924.
5. Jahanbin, Peyman. "Modeling L1 Influence on L2 Pronunciation: An MFCC-Based Framework for Explainable Machine Learning and Pedagogical Feedback." 2025, arXiv, arXiv:2504.13765.
6. Yaniafari, Rahmati Putri, Viga Olivia, and Suharyadi. "The Potential of ASR for Improving English Pronunciation: A Review." *KnE Social Sciences*, vol. 7, no. 7, 2024, doi:10.18502/kss.v7i7.10670.