DEVELOPMENT OF AN INTERACTIVE AI-BASED TOOL FOR AUTOMATIC PRONUNCIATION ASSESSMENT IN FOREIGN LANGUAGE LEARNING

Abdumalikov Akmaljon Abduxoliq o'g'li[1], Rohmonqulov Muhammadyusuf Egamberdi o'g'li[2]

[1]Jizzakh branch of national university of Uzbekistan, PhD, professor
[2]Jizzakh branch of national university of Uzbekistan, magistrant
Email: muhammadyusufrohmonqulov501@gmail.com
ORCID: 0009-0003-9976-0770

KEYWORDS

artificial intelligence, pronunciation assessment, speech recognition, adaptive learning, interactive web platform, CAPT systems

ABSTRACT

This paper presents the development of an AI-based interactive web platform designed for automatic pronunciation assessment and improvement in foreign language learning. The system analyzes learner speech using speech recognition and deep learning models, detects pronunciation errors, and provides adaptive feedback in Uzbek. The proposed framework integrates pronunciation accuracy, user engagement, adaptive learning mechanisms, and real-time feedback into a unified evaluation model. Experimental results demonstrate that regular use of the platform significantly improves pronunciation accuracy, intelligibility, and learner confidence. The findings confirm the effectiveness of AI-driven adaptive learning systems in enhancing foreign language pronunciation training.

INTRODUCTION

Accurate pronunciation plays a crucial role in effective communication in a foreign language. Even when learners possess sufficient vocabulary and grammar knowledge, poor pronunciation can reduce intelligibility and confidence. Adult learners often struggle to achieve correct pronunciation due to the influence of their native language phonetic system and limited exposure to native speakers [1]. With the rapid development of artificial intelligence, computer-assisted pronunciation training (CAPT) systems have become a promising solution for independent learning. These tools allow learners to practice repeatedly, receive instant feedback, and improve their speaking skills in a comfortable environment. Studies show that automated pronunciation training can significantly improve learners' intelligibility and speaking confidence. However, most existing systems have several limitations. They usually provide feedback only in English, which makes understanding errors difficult for beginners. Many platforms focus mainly on individual sounds and do not provide detailed explanations. In addition, adaptive and personalized learning features are often limited [2].

This research aims to develop an interactive web-based pronunciation learning tool designed specifically for Uzbek-speaking learners. The system uses AI technologies to evaluate pronunciation and provides explanations in Uzbek, helping learners better understand their mistakes and improve faster.

II. METHODOLOGY

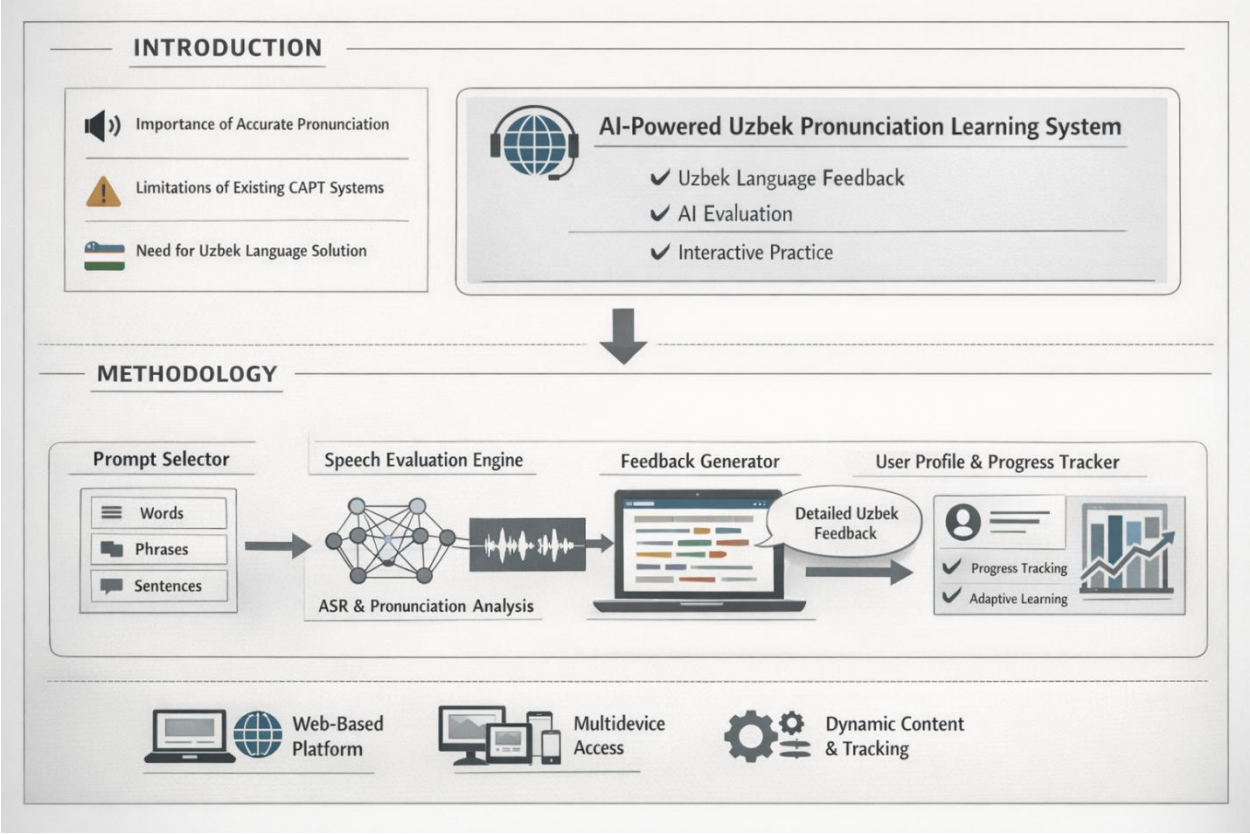
The proposed interactive pronunciation training system is implemented as an AI-driven web-based platform that integrates a speech processing backend with a user-centered frontend interface. The overall architecture consists of four primary components: a Prompt Selector, a Speech Evaluation Engine, a Feedback Generator, and a User Profile & Progress Tracker.

In a typical usage scenario, the system presents learners with pronunciation exercises in the form of words, phrases, or short sentences. The learner records their speech directly through a web

© A.A. Abdumalikov, M.E. Rohmonqulov

browser using a microphone. The recorded audio signal is transmitted to the backend server, where it undergoes automatic speech recognition and pronunciation analysis. The analysis results are processed by the Feedback Generator, which produces detailed corrective feedback in Uzbek. This feedback, together with performance scores and highlighted pronunciation errors, is displayed through the interface. Meanwhile, the User Profile module logs performance data, monitors progress over time, and dynamically adjusts the difficulty and content of subsequent exercises.

The web-based design ensures accessibility across multiple devices, including desktop computers, tablets, and smartphones, without requiring additional software installation. This architecture supports continuous system updates and simplifies the integration of additional modules or languages in the future. The backend is implemented using Python and modern machine learning frameworks, incorporating automatic speech recognition (ASR) models and custom pronunciation evaluation algorithms. The frontend interface is fully localized in Uzbek, including menus, instructions, and feedback messages, making the platform accessible even for users with limited English proficiency.



2.1-rasm. Architecture of an AI-based web system for evaluating pronunciation and providing feedback in Uzbek.

Automatic Pronunciation Assessment Engine. The central component of the system is the automatic pronunciation assessment engine, which analyzes learner speech to detect pronunciation errors and compute performance scores. A hybrid approach was adopted that combines established pronunciation scoring techniques with deep learning models to achieve both reliability and adaptability.

The evaluation process begins with speech recognition and forced alignment. The learner's recorded speech is aligned with the expected transcript of the prompt, allowing the system to determine the timing and likelihood values for each phoneme in the utterance. The system uses a pretrained acoustic model for English phonemes based on an HMM-DNN architecture trained on a large corpus of native English speech. Since the primary target users are Uzbek speakers, certain systematic pronunciation deviations are

© A.A. Abdumalikov, M.E. Rohmonqulov

anticipated, especially for sounds that do not exist in the Uzbek phonetic system, such as /θ/ and /ð/.

To improve detection accuracy, the acoustic model was fine-tuned using available samples of Uzbek-accented English speech. In addition, expected mispronunciation patterns were incorporated into the recognition lexicon. For example, substitutions such as /s/ for /θ/ were considered during alignment to enhance the model's sensitivity to common learner errors. This approach is conceptually related to the Extended Recognition Network technique, which integrates frequent mispronunciations into the recognition process to improve error detection [3].

Using the alignment and acoustic likelihood values, the system computes Goodness of Pronunciation (GOP) scores for each phoneme as well as an overall utterance score. GOP is calculated using the posterior probability of the target phoneme relative to alternative phoneme probabilities. Because a large-scale Uzbek learner speech corpus was not initially available, mispronunciations were simulated by deliberately modifying phoneme sequences in transcripts (e.g., replacing /r/ with /l/ or omitting final consonants) and analyzing the resulting score distributions. For each phoneme, decision thresholds were determined at the Equal Error Rate (EER), balancing false acceptances and false rejections. These calibrated thresholds allow the system to classify phonemes as correctly or incorrectly pronounced with reasonable accuracy [4].

To further enhance evaluation quality, a DNN-based classifier was integrated to assess pronunciation at the word level. This model uses multiple features, including phoneme-level GOP scores, acoustic confidence measures, and prosodic characteristics such as pitch and duration. The classifier was trained using a dataset of native and non-native speech samples labeled by expert evaluators. By combining rule-based GOP analysis with learned model predictions, the system produces both numerical scores (e.g., percentage ratings) and categorical feedback indicating which sounds were mispronounced. This hybrid framework enables a more comprehensive evaluation by capturing pronunciation aspects that GOP alone may not fully represent, such as rhythm and coarticulation effects.

Feedback Generation and User Interface. A distinctive characteristic of the system is the use of Uzbek as the primary language for feedback and instructional guidance. This design choice aims to reduce cognitive load and ensure that learners clearly understand the nature of their pronunciation errors. For each exercise, the Feedback Generator highlights specific words or phonemes that were mispronounced, provides an explanation in Uzbek, and suggests practical articulation strategies [5].

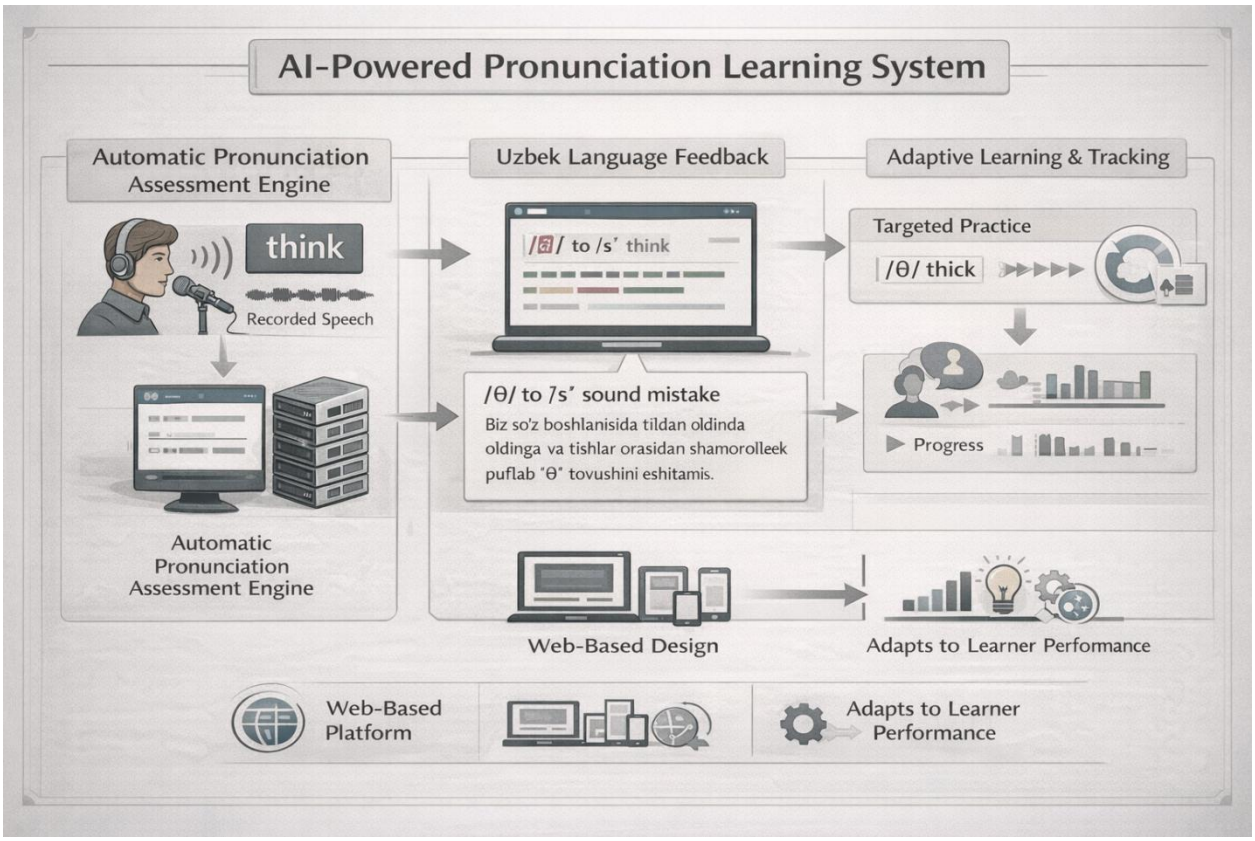
For example, if a learner pronounces "think" as "sink," the system identifies the incorrect /θ/ sound and provides guidance on proper tongue placement and airflow. Such native-language explanations make technical pronunciation concepts easier to understand, especially for beginners.

The interface also provides multimodal support through audio-visual feedback. Users can listen to native speaker pronunciations and compare them with their own recordings. Visualizations such as waveform and pitch contours allow learners to observe differences between their speech and the target model. Color-coded highlighting is used to indicate performance: mispronounced phonemes appear in red, while correctly pronounced segments are shown in green. An overall performance score is displayed after each attempt [6]. To increase engagement and motivation, the system allows repeated practice and tracks improvements over time. Gamification elements, such as achievement indicators and performance milestones, are included to encourage sustained use. Importantly, learners can practice at their own pace without social pressure, which supports confidence building and consistent repetition.

Adaptive Learning and Content Customization. The platform is designed to adapt to learners' proficiency levels and individual learning progress. New users are offered a short placement test consisting of basic pronunciation tasks [6]. Based on their performance, learners are categorized into general levels such as Beginner, Intermediate, or Advanced, and the content is tailored accordingly. As learners continue using the system, performance data is recorded and analyzed to identify recurring pronunciation difficulties. If consistent errors are detected in

(cc) BY © A.A. Abdumalikov, M.E. Rohmonqulov

specific phonemes (for example, /θ/ or /w/ vs /v/), additional targeted exercises are automatically introduced. Conversely, once certain sounds are mastered, more complex tasks are presented to maintain an appropriate level of challenge [7].



2.2-rasm. Process of analyzing learner speech and generating Uzbek feedback with progress tracking.

The system also adapts the level of feedback detail. Beginners receive more explicit explanations and step-by-step articulation guidance, while advanced users receive more concise corrective suggestions. Over time, this adaptive mechanism supports personalized learning and helps maintain learner engagement. Although the current implementation focuses on English pronunciation, the system architecture supports multilingual extensibility [8]. New acoustic models and pronunciation dictionaries can be integrated for other languages, allowing the platform to evolve into a broader pronunciation training system for multiple foreign languages.

III. RESULTS

To evaluate the effectiveness of the proposed pronunciation training tool, a series of experiments and a pilot user study were conducted to address two primary research questions: how accurately the automated assessment aligns with human expert evaluations, and whether regular use of the system leads to measurable improvement in learners' pronunciation.

Pronunciation Scoring Accuracy. To assess the reliability of the automated scoring, speech recordings were collected from 20 Uzbek learners of English representing different proficiency levels (approximately 10 beginners, 7 intermediate, and 3 advanced). Each participant was asked to read aloud 50 English sentences designed to cover a wide range of phonetic challenges, producing approximately 5–8 minutes of speech per learner [9]. All recordings were processed by the system to generate pronunciation scores and identify mispronounced sounds. For validation, two experienced ESL instructors independently evaluated the same recordings, rating pronunciation quality on a 10-point scale and noting significant pronunciation errors.

The comparison between system-generated scores and human ratings revealed a strong positive correlation, with a Pearson correlation coefficient of r = 0.82. This result indicates that the automated evaluation closely reflects human

© A.A. Abdumalikov, M.E. Rohmonqulov

perception of pronunciation quality. In addition, the mean absolute error (MAE) between the system's scores (scaled to a 10-point range) and the human ratings was 0.8, which represents a relatively low discrepancy for an automated assessment tool. These results are comparable to, and in some cases better than, earlier CAPT systems relying primarily on GOP-based scoring.

The system's ability to detect specific pronunciation errors was also examined. Based on expert annotations, 120 instances of notable pronunciation errors were identified, including dropped word-final consonants, /r/–/l/ confusion, and incorrect production of "th" sounds. The system successfully detected 102 of these cases, achieving a recall rate of approximately 85%. Around 15 false positives were observed, resulting in a precision rate of about 87%. Most false positives involved minor pronunciation variations that human raters considered acceptable but were flagged by the system as deviations. Overall, these findings demonstrate that the automatic evaluation engine is capable of reliably identifying pronunciation errors and producing scores that closely align with expert judgments. This level of agreement is critical for ensuring that learners can trust the system's feedback as an accurate reflection of their pronunciation performance.

Pronunciation Improvement in Learners. To evaluate the learning impact of the tool, a four-week pilot study was conducted with 10 Uzbek-speaking learners of English aged 18–30, representing beginner to intermediate proficiency levels. Participants were asked to use the system regularly, practicing for at least 30 minutes per day using the exercises provided. Before the training period, a pre-test was administered in which participants read a standardized English paragraph (approximately B1 level, containing diverse phonemes and multi-syllabic words). Recordings from this test were evaluated using the system's scoring metrics and also rated for intelligibility by a native English speaker on a 5-point scale. Specific recurring pronunciation errors were also documented for each participant.

After four weeks of practice, a post-test of similar difficulty was conducted using a different paragraph. The results showed clear and consistent improvement across all participants. The average pronunciation score increased by approximately 18 percentage points, rising from 62.4 in the pre-test to 80.7 in the post-test. Individual improvements ranged from +10 to +25 points. Human intelligibility ratings also improved significantly [10]. The average rating increased from 3.2 to 4.1 on a 5-point scale, and this change was statistically significant according to the Wilcoxon signed-rank test ($p < 0.01$). These findings indicate that improvements measured by the system were also reflected in human perception of speech clarity.

Qualitative observations revealed reductions in common pronunciation errors. Several learners who initially struggled with the /θ/ and /ð/ sounds demonstrated noticeable improvement in producing these sounds correctly. Another participant who previously misapplied word stress patterns showed better stress placement in the post-test, which contributed to improved intelligibility. Participants also reported positive learning experiences [11]. One learner noted that receiving immediate explanations in Uzbek helped them understand errors they had struggled with for years. This feedback supports the assumption that native-language explanations facilitate clearer comprehension of corrective guidance.

User feedback collected through questionnaires indicated that all participants found the system easy to use, and eight out of ten reported increased confidence in speaking English after the training period. Many appreciated the ability to practice privately and repeatedly at their own pace. While some users mentioned that the system could be strict in evaluating minor deviations, they also recognized that this encouraged more careful pronunciation practice.

IV. DISCUSSION

The strong correlation between the system's pronunciation scores and human expert evaluations ($r = 0.82$) represents a key outcome of this study. It confirms that the automatic assessment module can function as a reliable proxy for expert judgment in pronunciation evaluation. This level of accuracy is comparable to, and in some cases consistent with, state-of-the-art CAPT systems reported in the literature. Achieving such agreement is essential because learners must be able to trust the feedback

(cc) BY © A.A. Abdumalikov, M.E. Rohmonqulov

provided by the system for it to serve as an effective self-learning tool.

The combination of GOP-based phoneme analysis and DNN-based modeling appears to have played a crucial role in achieving this reliability. By integrating rule-based thresholds with data-driven learning, the system benefits from both structured linguistic knowledge and adaptive pattern recognition. This supports earlier research suggesting that deep learning methods enhance pronunciation scoring performance beyond traditional HMM-based approaches. Furthermore, the system's ability to detect specific pronunciation errors with approximately 85% recall demonstrates a strong diagnostic capability. Identifying precise articulation problems allows for targeted corrective practice rather than providing only general performance scores.

The presence of a small number of false positives reflects a known limitation of automated systems, which may interpret acceptable accent variations as errors. This highlights the need for a balanced tolerance level, particularly for advanced learners where complete native-like pronunciation may not be necessary. Future refinements could incorporate adaptive thresholds that account for acceptable accent variability while still maintaining pedagogical effectiveness.

The pilot study results also provide strong evidence of the system's practical value in pronunciation learning. An average improvement of approximately 18 percentage points in system scores, together with a significant increase in intelligibility ratings, indicates that consistent use of the tool can lead to meaningful pronunciation development within a relatively short time. These findings align with earlier CAPT studies reporting measurable gains in learners' speech clarity and confidence after regular automated practice. In particular, improvements in intelligibility, word stress, and sound accuracy suggest that the system supports both segmental and suprasegmental development [12]. This is an important outcome, as many traditional pronunciation tools focus primarily on individual phonemes. The continuous and immediate feedback provided by the platform likely plays a key role in reinforcing correct pronunciation habits. Unlike classroom settings, where correction may occur intermittently, the system provides feedback after every attempt, allowing learners to refine their pronunciation through repeated practice.

One of the most innovative aspects of the proposed system is the use of Uzbek-language explanations for corrective feedback. This feature addresses a significant limitation in many existing pronunciation training tools, which typically provide feedback only in the target language. User responses from the pilot study indicate that explanations in the native language improved understanding of pronunciation errors and reduced confusion [13]. By eliminating the need to interpret technical instructions in English, the system lowers cognitive load and allows learners to focus directly on correcting articulation. This approach may be especially beneficial for beginner and intermediate learners, who often struggle to understand phonetic explanations in a foreign language. From a pedagogical perspective, native-language feedback functions as a bridge between existing linguistic knowledge and new phonetic patterns. It enables clearer conceptual understanding of pronunciation mechanisms and may accelerate the acquisition of difficult sounds. This design choice also aligns with broader recommendations in educational technology literature that emphasize the importance of personalized and accessible feedback.

When compared with existing international CAPT solutions, several distinguishing features of the proposed system become evident. Many commercial pronunciation tools provide generic feedback without adapting to individual learning needs, whereas this system dynamically adjusts exercises based on user performance. Such adaptivity helps maintain an optimal level of challenge and supports more efficient learning. Additionally, while many studies emphasize segmental pronunciation training, the current system incorporates elements of prosodic feedback, including stress and pitch visualization, which appear to contribute to improved speech naturalness.

Another important contribution is the approach used to address limited training data. In the absence of a large Uzbek-accented speech corpus, simulated mispronunciations were used to calibrate scoring thresholds. This method provides a practical strategy for developing pronunciation assessment tools for under-resourced linguistic

(cc) BY © A.A. Abdumalikov, M.E. Rohmonqulov

contexts. Although performance could improve further with larger datasets, the results demonstrate that meaningful evaluation accuracy can still be achieved with carefully designed training strategies. Despite these promising outcomes, several limitations remain. The evaluation involved a relatively small number of participants and a short training duration, which restricts the generalizability of the findings. Larger and longer-term studies would be necessary to measure sustained learning effects and retention of pronunciation improvements. Additionally, while the system performs well in detecting segmental errors, its analysis of suprasegmental features such as rhythm, intonation patterns, and connected speech remains relatively basic. Future work could incorporate more advanced prosody analysis models to enhance this capability.

Another limitation is that the current system is designed primarily for read-aloud tasks, where the spoken content is known in advance. It does not yet support spontaneous speech evaluation or conversational interaction, which require more complex modeling of unpredictable language input. Integrating conversational AI technologies could enable real-time interactive speaking practice and expand the system's functionality into a full virtual pronunciation tutor. Looking ahead, the findings of this study open several directions for future development. Expanding the evaluation to larger learner populations and educational settings would provide stronger evidence of effectiveness. Enhancing feedback mechanisms with multimedia support, such as animated articulation diagrams, could further improve learner understanding. Additionally, incorporating more contextualized speaking activities, such as dialogue-based exercises, would allow users to practice pronunciation in communicative situations rather than isolated sentence reading.

V. CONCLUSION

In this paper, we presented the design, implementation, and evaluation of an AI-integrated interactive software tool for automatic pronunciation assessment and improvement, with a focus on Uzbek-speaking learners of English. By leveraging automatic speech recognition and machine learning, the tool can evaluate a learner's pronunciation with a high degree of agreement with human experts, and it offers suggestions for improvement in the learner's native language (Uzbek), which is a novel approach in the realm of pronunciation training. Our preliminary experiments confirmed that the tool's automated scores reliably reflect pronunciation quality and that regular practice with the tool led to marked improvements in learners' pronunciation accuracy, intelligibility, and confidence.

The scientific novelty of this work lies in its integration of advanced pronunciation assessment techniques with a user-centered, adaptive learning framework. We analyzed numerous international studies in CAPT and addressed some of their noted limitations – such as the lack of personalized feedback and the predominant focus on segmental drills – by incorporating personalized, L1-based feedback and including exercises on stress and intonation. Furthermore, unlike most existing CAPT systems, our tool is localized for a specific learner population, demonstrating how using the learners' native language in feedback can enhance the learning experience. This approach can serve as a model for other under-resourced languages and contexts, illustrating a path to bring cutting-edge language learning technology to a broader and more diverse audience.

From a practical perspective, the tool offers a scalable and accessible solution for pronunciation practice. It can be used independently by learners or integrated as a supplement in language courses. In regions where access to native speaker teachers is limited, such AI-driven tools can play a significant role in helping learners achieve better pronunciation, as evidenced by research and our own findings. The success of our system in a small pilot paves the way for larger deployments. We plan to continue improving the tool – expanding its capabilities to other languages and refining its feedback mechanisms – and to conduct broader evaluations. We will also explore collaborations with educational institutions to test the tool in classroom settings and gather more feedback.

REFERENCES

1. Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. Language Learning & Technology, 2(2), 62–76.

(cc) BY © A.A. Abdumalikov, M.E. Rohmonqulov

2. Witt, S. M., & Young, S. J. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. Speech Communication, 30(2–3), 95–108.
3. Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. Computer Assisted Language Learning, 15(5), 441–467.
4. Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners’ fluency by means of automatic speech recognition technology. Journal of the Acoustical Society of America, 107(2), 989–999.
5. Liakin, D., Cardoso, W., & Liakina, N. (2017). Learning L2 pronunciation with a mobile speech recognition app: French /y/. CALICO Journal, 34(1), 1–25.
6. McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. System, 57, 25–42.
7. Nodir, Rahimov, and Khasanov Dilmurod. "The mathematical essence of logistic regression for machine learning." International Journal of Contemporany Scientific and Technical Research 1.1 (2022): 102-5.
8. Dilmurod, H., and R. Shirinboyev. "THE MAIN THEORY OF LINEAR REGRESSION MODELS IN DATA MINING." Sun’iy Intellekt Nazariyasi va Amaliyoti: Tajribalar, Muammolar va Istiqbollari (2024): 143-147.
9. Rakhimov, N. O., D. R. Khasanov, and H. N. Abdulhakimov. "APPLICATION OF MACHINE LEARNING ALGORITHMS TO FORECAST THE YIELD OF GRAIN CROPS." Science innovation 3.Special Issue 59 (2024): 100-102.
10. Rohmonqulov , M. (2025). DEVELOPMENT OF AN INTERACTIVE SOFTWARE TOOL FOR AUTOMATIC ASSESSMENT AND IMPROVEMENT OF PRONUNCIATION IN FOREIGN LANGUAGE LEARNING. Scientific Practical Conference, 1(1), 344-349. https://d-pressa.com/index.php/spc/article/view/258
11. Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. Annual Review of Applied Linguistics, 27, 184–202.
12. Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In: Strange, W. (Ed.), Speech perception and linguistic experience. Timonium, MD: York Press.
13. Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine, 29(6), 82–97.

© A.A. Abdumalikov, M.E. Rohmonqulov