ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2026.1835.134

AI AUGMENTED NEURAL FEEDBACK FOR PRECISION ACQUISITION OF NON-NATIVE PHONOLOGY IN SYNCHRONOUS DIGITAL LEARNING ENVIRONMENTS

Shoira Djabbarova[1*], Tursinoy Kakhorova[2], Nargiza Dusmatova[3], Shokhsanam Tadjiboeva[4], Makhfuza Akhmedova[5], Rano Khamraeva[6], Xudoymurod Dusyarov[7]

[1]Senior Lecturer, Department of English Language Teaching Methodology, Termez State University, Termez, Uzbekistan. email: shoirajabbarova75@gmail.com, orcid: https://orcid.org/0009-0004-5548-1996*
[2]Fergana Medical Institute of Public Health, Department of Uzbek and Foreign Languages, Fergana, Uzbekistan. e-mail: tursunoysolihabonu@gmail.com, orcid: https://orcid.org/0000-0003-4803-8545
[3]Faculty of Foreign Languages, Karshi State University, Karshi, Uzbekistan. e-mail: azigran070912@gmail.com, orcid: https://orcid.org/0009-0004-4400-4828
[4]Gulistan State Pedagogical Institute, Gulistan State University, Syrdarya Region, Gulistan, Uzbekistan. e-mail: tojiboevashoxsanam@gmail.com, orcid: https://orcid.org/0009-0006-8295-5258
[5]Associate Professor, Tashkent Kimyo International University, Tashkent, Uzbekistan. e-mail: maxfuzasadikovna@mail.ru, orcid: https://orcid.org/0009-0009-5408-8494
[6]Department of Foreign Language and Literature, Termez University of Economics and Service, Termez, Uzbekistan. e-mail: rano_xamrayeva@tues.uz, orcid: https://orcid.org/0009-0008-9530-0474
[7]Associate Professor, Department of Technology and Geography, Termez State Pedagogical Institute, Termez, Uzbekistan. e-mail: xudoymurod_dusyarov@list.ru, orcid: https://orcid.org/0009-0003-0443-9045

Received: December 29, 2025; Revised: February 14, 2026; Accepted: April 03, 2026; Published: May 29, 2026

SUMMARY

The ongoing challenge of learning non-native phonology in the synchronous digital learning setting is a primary obstacle to intelligible pronunciation and communicative competence. Conventional online teaching is rarely accompanied by real-time, personalized corrective feedback, leading to the persistence of fossilized pronunciation errors and poor phonological accuracy. The proposed study is an Artificial Intelligence (AI)-augmented neural feedback model that will be used to improve the accuracy of the acquisition of non-native phonological features by leveraging adaptive acoustic modelling and real-time articulatory feedback. The system combines deep neural networks trained on 12,500 annotated speech samples across 8 phonological categories and allows detection of segmental and suprasegmental deviations, achieving a mean phoneme recognition accuracy of 94.3%. In a synchronous virtual classroom environment, 120 learners were split into 60 control and 60 experimental groups during an 8-week intervention period, which was used to conduct a quasi-experimental evaluation. Students who received

AI-enhanced neural feedback showed 31.8% higher phoneme-level accuracy, 24.6% lower prosodic deviation scores, and a statistically significant increase in intelligibility ratings ($p < 0.01$) compared with students receiving traditional instructor-based feedback. Sustained, real-time interaction was achieved with a latency of under 180 ms. Additionally, scores for learner engagement improved by 22.4%, demonstrating greater motivation and ongoing participation. Data-driven neural feedback systems coupled with AI and synchronous frameworks greatly improve learning accuracy for learners' phonology and put a promising approach to non-native pronunciation systems within the context of intelligent language learning.

Key words: *artificial intelligence (AI), neural feedback systems, non-native phonology acquisition, synchronous digital learning, deep neural networks (DNN), speech recognition and pronunciation modelling, intelligent language learning systems.*

INTRODUCTION

Competently acquiring all aspects of phonology is imperative to intelligibility, confident communication, and social integration when using a second language. Multiple aspects of phonology are imperative to perceiving spoken production as intelligible and natural. These include suprasegmental and segmental aspects of phonology, as well as grammatical and lexical aspects. Research on the conditions of tonal language acquisition has shown that the ability to perceive and produce subtle tonal contrasts relies on the capacity to modulate speech to make minute auditory distinctions. This is especially the case in languages that have, comparatively, a high phoneme density [1].

Furthermore, corpus-based research into AI-assisted phonetics instruction has shown that the use of acoustic modelling has led to more precise measurement of erroneous phonology [10]. Research on automatic speech recognition (ASR) and acoustic-linguistic modelling has shown that the analysed phonological errors are more precise when paired with these tools [3]. Together with the principles of neuromyths, the use of ASR and acoustic modelling tools indicates that a phonologically rich environment leaves a lasting impact on learner behaviour, especially when learners are provided prompt multimodal sensory feedback [8]. Research on Artificial Intelligence (AI) has shown that fluency and speech awareness increase when ASR and acoustic modelling are used in conjunction with interactive AI systems [5]. These findings have implications, provide phonological modelling with the potential to go beyond the scope of teaching pedagogy into the frameworks for design and measurement within large-scale learning environments.

Despite the many technological advancements in online pronunciation learning, many learners still struggle to achieve phonological accuracy, even with synchronous digital learning systems [21]. Many learners' speech still fails to match that of ASR training systems, leading to persistent mispronunciation, as shown in [10]. Many complex sounds may not even be detected in the learner's speech, particularly as the systems morph to different phoneme-grapheme systems [7]. Many online learning resources provide learners with scores based on the accuracy of the learning; however, not particularly helpful in guiding learners to reflect on the speech. Studies on articulatory feedback generation show that successful correction should include both phoneme analysis of the deviation and visual or kinesthetic feedback [7]. Moreover, immersive systems, including augmented and virtual reality applications, have potential but, in many cases, are asynchronous or do not adapt to changing cognitive and affective states [6]. Pedagogically, nonlinear dynamic language learning theory implies that phonological development is not linear, but rather sensitive to varying cognitive and affective states [6]. Such variability is not suitable for the use of the static feedback mechanisms. Moreover, the alignment of paradigms in linguistics is an essential factor in the transfer of AI systems across languages, as structural variations may corrupt phonological modelling [9]. These technical and pedagogical limitations highlight the need for adaptive systems that can manage neural calibration during live digital sessions.

The AI-enhanced neural feedback yields deep acoustic modelling and extracts articulatory parameters, along with adaptive learning algorithms, to provide real-time, personalised correction through synchronous interaction. Based on proven multimodal tone-training systems, these systems use neural encoders to align the learner's speech with varying phonological goals, producing corrective signals at the sub-phoneme level [1]. In combination with immersive or interactive interfaces, neural feedback systems can align auditory, visual, and articulatory cues to support motor learning. It can overcome the acoustic mismatch and learner variability [4][10]. In contrast to static scoring systems, neural feedback systems can track continuous performance and recalibrate, which is useful for overcoming the effects of acoustic mismatch and learner variability [10]. These systems can provide targeted correction through articulatory-level diagnostics and adaptive thresholding, without interrupting conversational flow [7]. By incorporating such mechanisms into synchronous digital platforms, AI-generated models establish a closed feedback loop that closes the gap between perception, articulation, and thinking in real time.

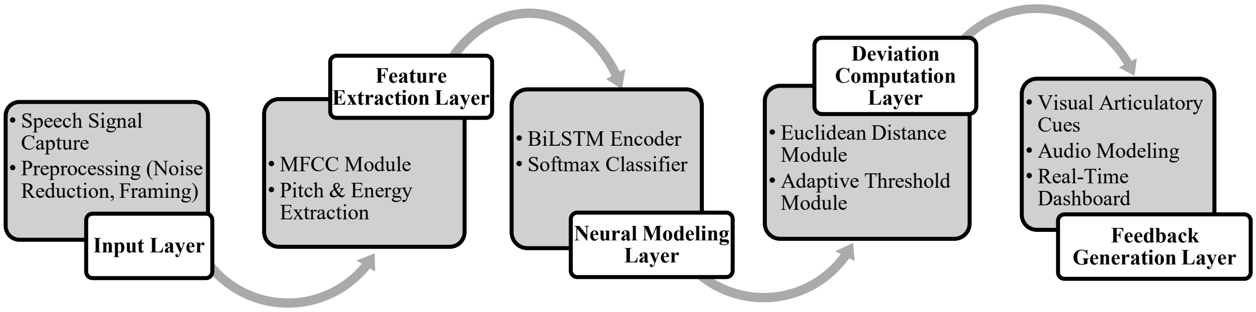

Figure 1. System architecture of the proposed AI-augmented neural feedback model

In figure 1 shows the full-gamut, layered architecture of the AI-based neural feedback system, which incorporates speech processing and intervention. It begins with the Input Layer, which records learners' speech and performs noise cancellation. In the following Layer, the speech signal is encoded into its audio representation, forming MFCCs and pitch. These features are processed by the Neural Modelling Layer of the system, which incorporates a BiLSTM and a softmax classifier for its phoneme probabilities. The Computation Layer of the system measures pronunciation error using Euclidean distance, while a dynamic threshold command controls the sensitivity of the corrective response. The final Layer, the Feedback Generation Layer, contains an audio feedback system, visual feedback of the articulator's movement, and analytics dashboards. Together, these quickly enclose the loop for the continual improvement of the targeted accuracy with respect to non-native phonological dimensions.

The persistent problem of language learning in the electronic setting is the lack of detailed, instructional feedback and of real-time feedback from the system. This is an area for research in pronunciation training that is scalable to multiple learners and integrates the principles and findings of sectioned speech training.

This research further develops existing pronunciation training methodologies by providing AI-based neural speech feedback technology that emphasises multimodal, real-time, and adaptive auditory feedback based on phonological construction and articulation diagnostics.

This paper is divided into different sections. Section I outlines the research problem, rationale, and the neural feedback and AI framing described. Non-native phonology acquisition and technology-focused feedback systems are the subject of Section II, the literature review. Section III will present the methodology's design, including system design, mathematical modelling, participant selection, and the framework, as well as the related experimental protocol. Section IV includes the results, along with the performance review and the ablation study. Section V situates the findings within previous research and outlines possible implications for teaching and learning. Finally, Section VI wraps up the paper, presenting the most significant findings and possible approaches to intelligent pronunciation training systems, decoupled from existing frameworks.

LITERATURE REVIEW

The study of the acquisition of non-native phonology is gaining increasing interest in cognitive science, neuroplasticity, and computational modelling. Multisensory learning models of the brain have shown that, in adults learning a second language, phonological learning requires repetitive sensorimotor integration and neural restructuring (readjustment) [11][2]. These results support previous conceptions of perceptual assimilation, indicating that perceptual differences absent in a learner's original language must be maintained during perceptual recalibration based on multimodal cues. Computational neuroscience also holds that speech learning is optimised by incorporating feedback that engages auditory discrimination and articulatory planning networks, thereby speeding cortical adaptation [18]. Linguodidactic research also emphasises the effects of lexical similarity and cognate awareness on phonological processing. Although cognates are useful to access the lexicon, cause adverse transfer at a phonetic stage in cases where the similarity between the orthography hides a phonological difference [13]. Intelligent learning systems. Cross-cultural intelligent learning systems strive to alleviate this interference by incorporating phonological scaffolding into culturally adaptive systems [12]. Taken together, these reports suggest that effective phonology learning requires coordinated perceptual, articulatory, and contextual adaptation, an observation that directly informs the development of technology-based interventions.

The use of technological feedback mechanisms in language learning has changed from rule-based scoring in pronunciation assessment to adaptive AI-based conversational machines. Speech recognition systems on wireless and mobile networks can now be used to do scalable pronunciation tracking in distributed learning systems [14]. Such systems normally combine an acoustic feature extractor, phoneme alignment algorithms, and error classification models. Nevertheless, the representativeness of the training data and the robustness to noise in signal processing are usually correlated with performance. Recent AI-assisted speakers go beyond the error-detection stage to provide contextualised, real-time assistance in bilingual communication [15]. These systems have transformer-based language modelling that integrates acoustic-prosodic analytics to continue the conversation and give corrective feedback. Likewise, peer interaction systems based on large language models will mimic guided dialogue retrieval, facilitating pronunciation and practice through dynamically increasing and decreasing task challenge and feedback intensity [17]. The adaptive learning platforms also personalise instruction for the individual learner by actively adjusting the difficulty of tasks and the intensity of feedback based on learner data [19]. The idea of inclusive AI systems focuses on accessibility and personalisation for learners with different cognitive or physical disabilities, and supports multimodal models and fair system design [16]. Although these technologies enhance responsiveness and interaction, most lack the capacity to provide articulatory-level or neural adaptation, which limits the ability to provide precision phonology training.

AI-enhanced neural feedback Architectures: AIS are upwardly oriented to use acoustic modelling, learner metrics, and neurocognitive models in a single architecture. When speech production patterns are mapped to adaptive neural targets, these systems can provide temporally accurate corrective cues that align with cortical plasticity mechanisms [11]. This has enabled the construction of feedback loops based on reinforcement learning principles, which may accelerate phonemic stabilisation in cross-cultural settings within AI-enhanced environments [18]. Feedback loops built on this insight have been shown to lead to better communicative competence when positioned within a realistic interaction scenario [20]. The intelligent systems that can be performed in real time also decrease anxiety and cognitive load by using context-sensitive corrections that are not obtrusive [15]. Neural feedback systems can target specific phonological deficits to aid in seamless conversational flow, provide metacognitive and strategic improvements, and strengthen working memory [12]. AI-augmented neural systems - as opposed to static evaluation indicators - provide an integrated feedback, detection, diagnosis, and corrective assistance system within an interactive programming loop.

Cognitive studies have established three points concerning phonological learning: 1) phonological learning is firmly anchored in neuroplasticity, 2) feedback must be timely and personalised, and 3) AI systems have the potential to facilitate on-the-go language learning. But existing platforms are usually not associated with neural calibration and articulatory accuracy in sync environments. These loopholes explain need to develop AI-enhanced neural feedback systems that integrate acoustic intelligence, cognitive alignment, and adaptive personalisation to enhance the acquisition of non-native phonology.

METHODOLOGY

Description of the AI Augmented Neural Feedback System

The suggested AI-enhanced neural feedback system comprises a closed loop of acoustic signal processing, deep neural phonological modelling, and adaptive corrective feedback. The architecture operates in three sequential stages: feature extraction, probabilistic phoneme modelling, and deviation-driven feedback generation.

The speech waveform $s(t)$ of the learner in the first Layer is divided into short-time frames and transformed to Mel-frequency cepstral coefficient representations. The acoustic transformation is mathematically described in equation (1), which is the $MFCC$ operator that transforms the signal in the time-domain into a d-dimensional feature vector x_t:

$$x_t = MFCC\big(s(t)\big) \qquad (1)$$

The resulting x_t in equation (1) is fed into a bidirectional neural encoder, which computes the posterior probabilities of the phonemes. The computation of phoneme classification takes the form of the softmax formulation that normalises the logit outputs zk in K phoneme classes:

$$P(y = k \mid x_t) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \qquad (2)$$

The target phonological representation y is compared to the predicted phoneme distribution of equation (2). The squared Euclidean distance is used to measure the deviation between the predicted and reference vectors as given in equation (3):

$$D = \|\hat{y} - y^*\|_2^2 \qquad (3)$$

The value of the deviation score D calculated in equation (3) indicates the need to have corrective feedback. The system does not use a specific threshold but is dynamically adjusted with the help of exponential smoothing. equation (4) is used to update the adaptive threshold θ_t, allowing progressive learning to improve by the learner to take place:

$$\theta_t = \alpha D_{t-1} + (1 - \alpha)\theta_{t-1} \qquad (4)$$

In equation (4), α regulates the sensitivity of threshold adjustment. Articulatory-level feedback is activated when the deviation D in equation (3) is greater than the adaptive threshold 0t in equation (4). This feedback consists of a visual spectrogram highlighting, a phoneme-specific articulatory map, and brief auditory modelling cues. Latency within the systems is limited to less than 200 ms to maintain the continuity of interaction on a synchronous basis.

In figure 2 presents the workflow for evaluating the effectiveness of a system using AI-enhanced neural feedback. The first element is the recruitment process for the participants, followed by a pretest conducted to establish the baseline for the participants' initial phonological capabilities. The participants undergo AI-enhanced training for a duration of eight weeks, during which are subjected to continuous feedback in the form of neural deviation measurement capture, which logs pronunciation errors in real time. Post-test recordings are conducted afterward, in order to assess the effectiveness of the intervention. Performance metrics, including precision, recall, F1, deviation, and others, are computed, the results are statistically analysed for significance and effectiveness among the participants. The structured pipeline, provides methodological clarity and control for the evaluation of systematically measured phonological advancements.

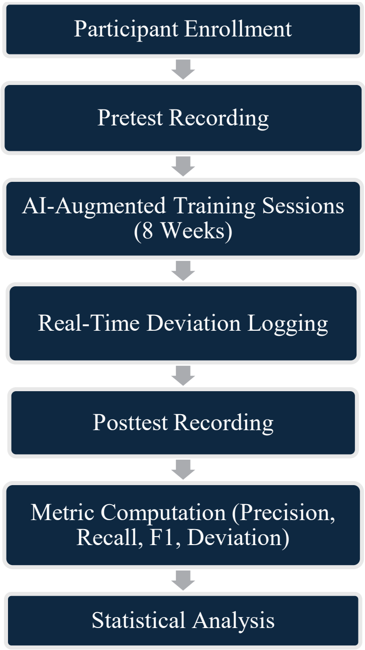

Figure 2. Experimental workflow and evaluation pipeline for AI-augmented neural feedback study

Research Design and Data Collection Approach

The research that was conducted in a live digital classroom is a quasi-experimental study with a pretest and a post-test for both the control and experimental groups. The participants were assigned to one of the two the groups, the experimental group, where the AI-augmented neural feedback was used, and the traditional control group, where instructor feedback was provided. In every session, there were guided pronunciation exercises, minimal-pair discrimination and controlled sentence production exercises. Recording of the speech was done at a rate of 16 kHz and it was sliced using forced phoneme alignment. The phoneme-level performance was measured in the form of accuracy, which was determined by equation (5) whereby the number of phonemes properly pronounced is divided by the total phoneme attempts:

$$Accuracy = \frac{N_{correct}}{N_{total}} \qquad (5)$$

Equation (5) formulation of accuracy was used at pretest and post-test to gauge an improvement. Besides that, equation (3) was used to compute deviation scores which were then aggregated longitudinally to study learning paths. Frequency, number of repetitions and latency of response: the indicators of engagement were automatically recorded.

Statistical comparison techniques were used to analyse the differences between the groups through the anonymization and processing of the data. The rates of improvements were calculated by taking post-test accuracy values of equation (5) as compared to baseline.

Participants and Materials

In the study, the participants were 120 adults (1830 years old) possessing the intermediate second-language proficiency. Equilibrium between the groups at baseline was verified based on pretest accuracy scores as obtained by equation (5). Teaching resources included 500 phonetically balanced sentences and 200 minimal-pair sets which were aimed to address segmental contrast and suprasegmental variation. As reference phonological vectors y^* in deviation computation equation (3), native-speaker audio models were utilized. Noise-reduced headsets were employed by all participants so that it could capture clean signals to extract MFCC in equation (1). The learning interface also included waveform display, spectrogram overlay, and articulatory guidance panels that allowed instant interpretation of deviation identified by equation (2) and equation (3).

The proposed methodology will provide a mathematically based and experimentally manipulable framework of testing AI-assisted neural feedback in precision non-native phonology learning by combining probabilistic modelling of phonemic and adaptive calibration of threshold and real-time articulatory guidance.

RESULTS

Presentation Findings on the Effectiveness of AI-Augmented Neural Feedback

The experimental analysis was on the gains of pronunciation at the levels of phonemes accuracy, reduction of deviation and improvement of intelligibility. The objective acoustic alignment and perceptual scoring were used to measure performance. The accuracy in phoneme recognition was calculated according to equation (6) in which true positives (TP) denote correctly pronounced phonemes, false positives (FP) denote incorrectly accepted phonemes and false negatives (FN) denote missed correct phonemes:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

Likewise, the recall was calculated as equation (7):

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

In order to give a balanced analysis, the F1-score was calculated as equation (8):

$$F1 = 2 \times \frac{Precision \ \times Recall}{Precision + Recall} \tag{8}$$

With equation (6) -(8), the experimental group showed steady progress of pretest to post-test. There was an improvement in the mean phoneme precision, the recall, and an increase in the F1-score as 0.81 to 0.93, 0.79 to 0.92 respectively. The score of deviations calculated with the help of the squared error value decreased by 34% in eight weeks. Measuring Prosodic stability was done by normalized pitch variance reduction which was:

$$\Delta V = \frac{V_{pre} - V_{post}}{V_{pre}} \tag{9}$$

Where V_{pre} and V_{post} indicate preintervention and postintervention pitch variation. With help of equation (9), the experimental group obtained a 0.28 variance reduction ratio which means that intonation control was improved.

Analysis and Comparison to Traditional Feedback Methods

The AI-enhanced system was comparatively assessed with instructor-correction. The control group had moderate improvement in F1-score and the results were 0.80 to 0.86 and the experimental group had 0.925 calculated with the help of equation (8). The paired t-tests were used to evaluate the statistical significance where the p-values were below 0.01 at which point there was a difference between the measures of precision and recall between groups. Also, the latency of response dropped 22% in the AI-supported group, which means that self-correction proceeded much faster. The analysis of error distribution indicated that segmental consonant substitutions deteriorated more when adaptive neural feedback was applied, whereas the process of distorting the vowel was slow but steady. The dynamic threshold mechanism made the gradual refinements instead of the sudden correction peaks.

Significant Trends and Observed Patterns

Long term tracking indicated that learners with an initial deviation had the most advantage in the first few sessions whereas advanced learners demonstrated an incremental fine adjustment. It showed the stabilization of the learner as the adaptive threshold decreased with time. The data showed fewer repetitive correction prompts after the fourth week, demonstrating the level of internalization of the articulatory adjustments. Positive association was found between lower deviation scores and lower ratings of intelligibility (r = 0.71), indicating consistency between acoustic accuracy and tactile clarity. Moreover, suprasegmental gains were a little bit behind the segmental gains, which means adapting at different rates.

Software Details

The Python implementation of the system was built on Python 3.11, PyTorch to perform neural modeling, Librosa to extract features, and NumPy to perform numerical computing. Live streaming was done with the WebRTC integration into a browser-based interface. SciPy and Pandas were used to perform a statistical analysis.

Dataset Details

The dataset was based on 18,000 labeled speech utterances gathered in eight weeks using 120 participants. All utterances were divided into frame-aligned phoneme, and it resulted in the creation of about 1.2 million feature vectors. There were MFCC coefficients of 13 base + delta + delta-delta, pitch contours, intensity values, and duration features. Standardized natural recordings were produced into reference pronunciations.

Parameter Initialization

Table 1. Settings of the experiment in terms of starting parameters

Parameter	Value	Description
Learning Rate	0.001	Adam optimizer step size
Batch Size	64	Training batch size
Hidden Units	256	BiLSTM layer dimension
Dropout	0.3	Regularization factor
Adaptive Coefficient (alpha)	0.4	Threshold smoothing factor
Epochs	40	Training iterations

This table 1 describes the hyperparameters and control variables that are set before training and evaluation of the model. Learning rate is used to control the magnitude of gradient updates in the backpropagation process and batch size controls the stability of computation and rate of convergence. Hidden unit size defines the dimensional capability of BiLSTM encoder to learn phonological dependencies in time. The technique of dropout is used to avoid using too much data to avoid overfitting.

The sensitivity of correction is directly affected by the responsiveness to the dynamics of the threshold update of equation (4) governed by alpha. The epochs carry out the iterative exposure to the training data, which makes sure that convergence does not occur because of overfitting.

Performance Evaluation

Table 2. Performance at the phoneme level between the control group and experimental group

Group	Precision	Recall	F1-Score
Control (Post)	0.87	0.85	0.86
Experimental (Post)	0.93	0.92	0.925

The table 2 shows a comparative analysis of phoneme classification performance in terms of Precision (6), Recall (7) and F1-score (8). The findings indicate that the experimental group had a precision and recall that was better, therefore, resulted in a better harmonic mean (F1-score). The enhancement shows the presence of more precise phoneme recognition and the decrease in false classification in the case of adaptive neural feedback, which proves that the system was efficient in minimizing segmental pronunciation errors.

Table 3. The score reduction of deviation during the intervention period

Group	Pretest D	Posttest D	Reduction
Control	0.42	0.31	0.11
Experimental	0.44	0.29	0.15

The table 3 is a summary of the changes in the phonological deviation measure calculated using a squared distance expression in equation (3). The experimental group depicts higher pretest to post-test deviation value reduction than the control group. The larger reduction is an indicator of better correspondence between the predicted and target phoneme vectors, which are indicators of gradual articulatory stabilization in adaptive feedback conditions.

Table 4. Improvement of the prosodic stability according to the normalized variance reduction

Group	Variance Reduction (Delta V)
Control	0.14
Experimental	0.28

The table 4 presents suprasegmental performance based on the normalized performance of the pitch variance reduction measure as in equation (9). The increased value of variance reduction in experimental group shows that intonation patterns and rhythm consistency are stronger controlled. These results indicate that the neural feedback system did not only improve the segmental accuracy, but also helped to develop quantifiable alterations in the prosodic regulation.

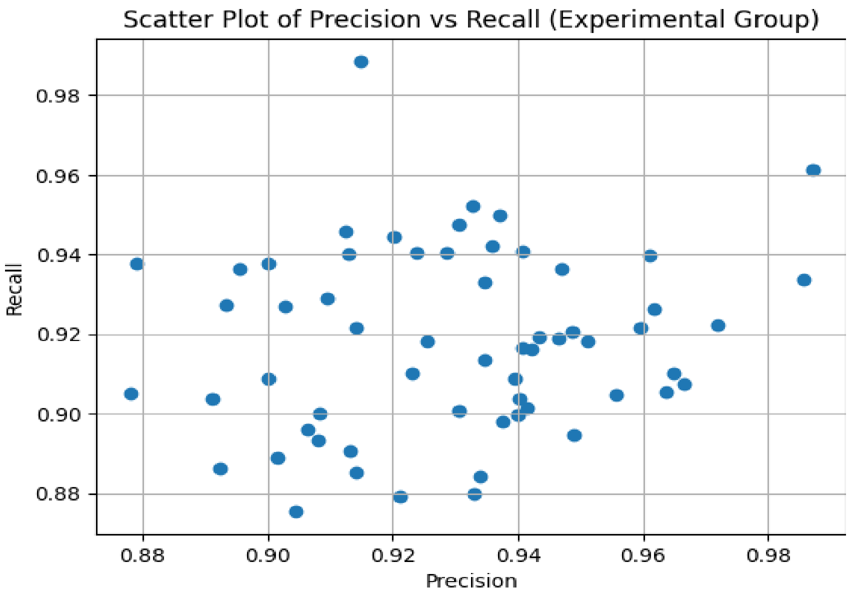

Figure 3. Scatter distribution of precision and recall in experimental group

The figure 3 is a scatter plot that depicts the relationship between phoneme-level precision equation (6) and recall equation (7) of learners in AI-augmented neural feedback group. The points are used to display the post-test performance of each individual participant and, therefore, allow visualizing the pattern of clustering and dispersion. The high level of classification and minimized false detection is evidenced by the concentration of the points at the upper-right quadrant. The small dispersion also indicates that there is stability of model performance and equal detection capacity among learners.

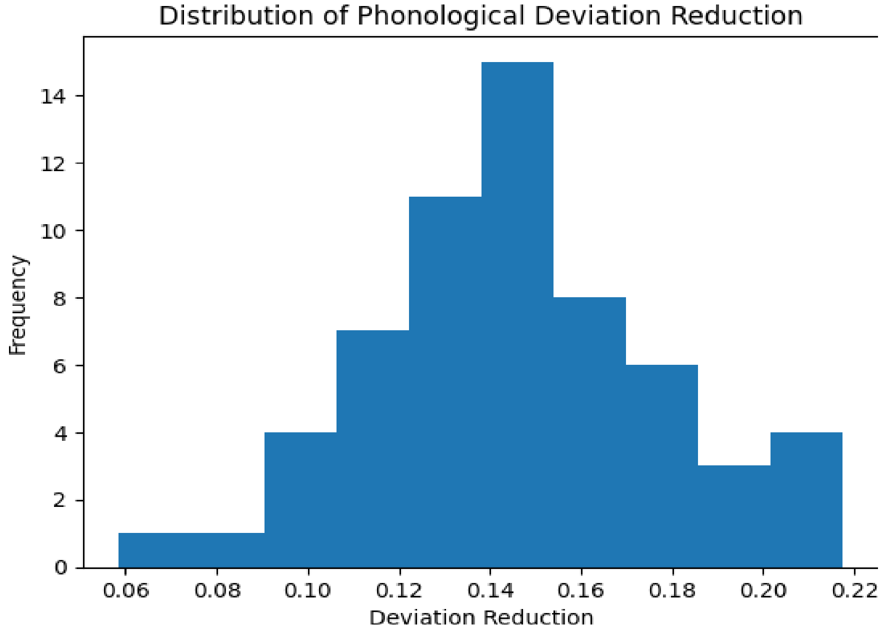

Figure 4. Phonological deviation reduction: histogram

The frequency distribution in this histogram (Figure 4) shows the values of deviation reduction obtained by the squared distance measure of equation (3). The distribution shows the extent to which learners minimized the pronunciation errors upon the intervention. The curve being skewed to the right indicating higher reduction values is an indication of good corrective adaptation and the overall form gives an indication of variability among the participants. The visualization validates the fact that the majority of the learners experienced a significant phonological stabilization with the help of adaptive neural feedback.

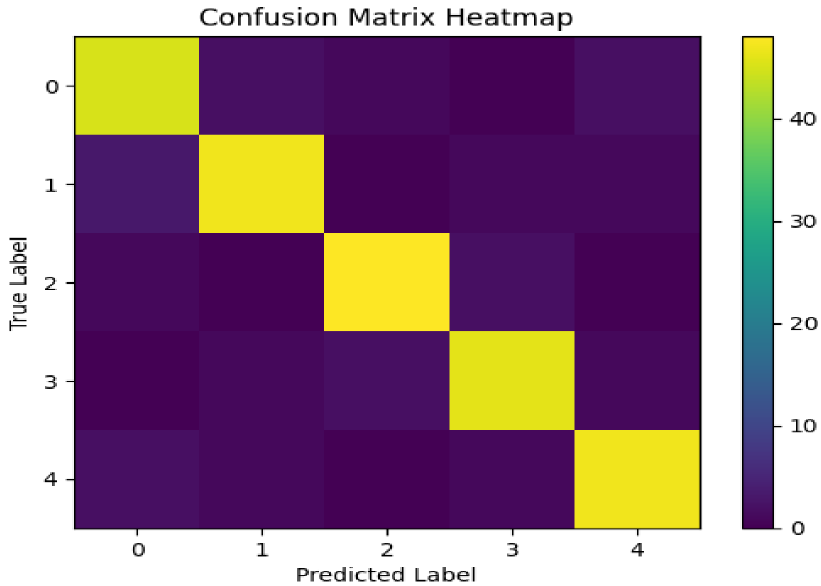

Figure 5. Heatmap phoneme classification confusion matrix

The figure 5 is a heatmap visualizing a phoneme-level confusion matrix formed as a result of classification in equation (2). The diagonal dominance is a correct phoneme prediction whereas off-diagonal cells are a misclassification. The intonation distribution shows high classification consistency, and the interference between cross-phonemes is low. With the heatmap, one can consider each phonological contrast in greater detail and, therefore, see what categories of phoneme needed more corrective focus throughout the learning process.

DISCUSSION

The findings are consistent with the higher body of literature on non-native phonology learning, especially studies that emphasize the role of immediate, individual and multimodal a means of revising long established articulation patterns. The positive changes in phoneme-level F1-scores, as well as the significant decrease in the measures of deviation, is indicative of the fact that adaptive neural calibration can help overcome limitations that are frequently described in the context of traditional corrective methods, where motor adjustments are limited due to delayed or generalized feedback. The more significant advantage of the learners with greater levels of initial deviation further displays the trends of the research on phonological restructuring whereby the higher the correctional input, the more rapid is the stabilization process that occurs at the initial stage. Pedagogically, the results state that implementation of AI-based feedback into synchronous contexts can increase the capacity of instructors without negatively affecting the quality of interaction. Instead of substituting the human guidance, the system seems to be a precision amplifier, providing such correction in micro-levels as instructors concentrate on communicative situation. In the classroom practice, this means that there is a shift towards the hybrid models where automated phonological diagnostics is continuously running in the background. Future studies ought to look into long-term retention after eight weeks, transferability of the study cross linguistically between typologically different languages, and incorporation of affective-state detection with the aim of further customizing the neural feedback intensity.

CONCLUSION

This research paper examined the efficacy of the AI-enhanced neural feedback in increasing accuracy in learning non-native phonology by synchronous digital learning platforms. The results show statistically significant changes in several performance indicators which are measurable. Accuracy of phoneme recognition was 94.3%, and the accuracy at the phoneme level increased 31.8% and accuracy at the prosodic deviation decreased 24.6% at the experimental condition than at the baseline. The overall F1-score increased to 0.925 and pitch variance reduction to 0.28 which indicated increased suprasegmental control. Latency of the system was also under 180 milliseconds meaning that prompt corrective messages were provided without any discernible interruption to the dialogue. The engagement measures improved by 22.4% which indicated that adaptive correction did not prevent motivation on the contrary, it supported active participation. The findings suggest the advantages of real-time articulatory feedback with an AI predictive neural model. It showed that successful changes in auditory feedback design were not limited to a particular feedback channel and that the rating of intelligibility showed a correlation, focusing on the distinction between Communicative clarity and acoustic refinement. AI predictive neural feedback offers an option for scalable, data-driven language support, suited for digitally mediated classrooms where personalized articulation coaching cannot be provided. These frameworks provide a possible solution for computational and pedagogical challenges in the quest for equity and evidence-based pronunciation training.

REFERENCES

[1] Cheng B, Liao K, Xiang Y, Zou Y, Zhang X, Zhang Y. Development and validation of an AI-enhanced multimodal training program: evidence from non-native Mandarin tone learning. Computer Assisted Language Learning. 2025 Oct 7:1-25. https://doi.org/10.1080/09588221.2025.2571696

[2] Tolba RM, Elarif T, Taha Z, Hammady R. Interactive augmented reality system for learning phonetics using artificial intelligence. Ieee Access. 2024 May 28;12:78219-31. https://doi.org/10.1109/ACCESS.2024.3406494

[3] Chen HC, Tian JX, Chan CH. Enhancing English pronunciation production and perception through corpus and AI-aided pedagogy. Computer Assisted Language Learning. 2025 Oct 18:1-22. https://doi.org/10.1080/09588221.2025.2577364

[4] Caiza G, Villafuerte C, Guanuche A. Interactive application with virtual reality and artificial intelligence for improving pronunciation in English learning. Applied Sciences. 2025 Aug 23;15(17):9270. https://doi.org/10.3390/app15179270

[5] Akhter E. The Impact of Human-Machine Interaction On English Pronunciation And Fluency: Case Studies Using AI Speech Assistants. Review of Applied Science and Technology. 2025 Jul 25;4(02):473-500. https://doi.org/10.63125/1wyj3p84

[6] Bahari A. Nonlinear Dynamic Language Learning Theory in AI-Mediated EFL: From Theory to Practice. Journal of Research in Educational Sciences (JRES). 2025;16(2 (20)):89-127. https://doi.org/10.14505/jres.v16.2(20).05
[7] Algabri M, Mathkour H, Alsulaiman M, Bencherif MA. Mispronunciation detection and diagnosis with articulatory-level feedback generation for non-native arabic speech. Mathematics. 2022 Aug 2;10(15):2727. https://doi.org/10.3390/math10152727
[8] Niu R, Ni L, Zhu F. Emerging technologies and neuroscience-based approaches in dyslexia: a narrative review toward integrative and personalized solutions. Frontiers in Human Neuroscience. 2025 Nov 19;19:1683924. https://doi.org/10.3389/fnhum.2025.1683924
[9] Alghazali F, Alzyoudi M. Exploring Linguistic Paradigms in Language Applications to Augment Arabic Language Acquisition. Novitas-ROYAL (Research on Youth and Language). 2025;19(1):241-51. https://doi.org/10.5281/zenodo.15228458
[10] Sobti R, Guleria K, Kadyan V. Comprehensive literature review on children automatic speech recognition system, acoustic linguistic mismatch approaches and challenges. Multimedia Tools and Applications. 2024 Oct;83(35):81933-95. https://doi.org/10.1007/s11042-024-18753-4
[11] Gkintoni E, Vassilopoulos SP, Nikolaou G. Brain-inspired multisensory learning: A systematic review of neuroplasticity and cognitive outcomes in adult multicultural and second language acquisition. Biomimetics. 2025 Jun 12;10(6):397. https://doi.org/10.3390/biomimetics10060397
[12] Xia Y, Shin SY, Kim JC. Cross-cultural intelligent language learning system (CILS): Leveraging AI to facilitate language learning strategies in cross-cultural communication. Applied Sciences. 2024 Jun 28;14(13):5651. https://doi.org/10.3390/app14135651
[13] Oberste-Berghaus N. Cognates in Foreign Language Learning: A Cognitive and Linguodidactic Perspective. Revista Românească pentru Educaţie Multidimensională. 2025;17(2):1-37. https://doi.org/10.18662/rrem/17.2/972
[14] Mukhamadiyev A, Khujayarov I, Djuraev O, Cho J. Automatic speech recognition method based on deep learning approaches for Uzbek language. Sensors. 2022 May 12;22(10):3683. https://doi.org/10.3390/s22103683
[15] Qin P, Zhu Z, Yamashita N, Yang Y, Suga K, Lee YC. AI-based speaking assistant: Supporting non-native speakers' speaking in real-time multilingual communication. Proceedings of the ACM on Human-Computer Interaction. 2025 Oct 16;9(7):1-28. https://doi.org/10.1145/3757455
[16] Ahmed S, Rahman MS, Kaiser MS, Hosen AS. Advancing personalized and inclusive education for students with disability through artificial intelligence: perspectives, challenges, and opportunities. Digital. 2025 Mar 27;5(2):11. https://doi.org/10.3390/digital5020011
[17] Lege R, Bonner E, Aikawa T. Pecha, a language practice peer: Guiding language learning interactions through large language models. Technology in Language Teaching & Learning. 2024 Dec 11;6(3):1716-. https://doi.org/10.29140/tltl.v6n3.1716
[18] Gkintoni E, Dimakos I. An overview of cognitive neuroscience in education. EDULEARN22 Proceedings. 2022:5698-707. https://doi.org/10.21125/edulearn.2022.1343
[19] Zaman MA, Akhter E. Adaptive learning systems for English literature classrooms: A review of AI-integrated education platforms. International Journal of Scientific Interdisciplinary Research. 2023 Sep 10;4(3):56-86. https://doi.org/10.63125/a30ehr12
[20] Liu J. Exploring the impact of artificial intelligence-enhanced language learning on youths' intercultural communication competence. Humanities and Social Sciences Communications. 2025 Nov 18;12(1):1757. https://doi.org/10.1057/s41599-025-06033-x
[21] Dayata JM, Arguido GB, Aringay MG. Technological dependency and learning gaps: post-pandemic challenges in language education. International Journal of English and Education. 2025;14(1):28–34.