Evaluating the Efficacy of a Mobile Augmented Reality Application for Enhancing Pronunciation in Uzbek Preschool Children

Gulnora Rakhimbaeva and Zarrina Kosimova
Department of Neurology and Medical Psychology, Tashkent State Medical University, Farobiy Str. 2, 100109 Tashkent, Uzbekistan
zarrinael@gmail.com, gulnora.rakhimbayeva@mail.ru

Keywords: Augmented Reality, Mobile-Assisted Language Learning, Early Childhood Development, Speech Sound Production, Uzbek Phonetics, Educational Technology, Preschool Education, Articulation Therapy, Digital Pedagogy.

Abstract: This study assessed the impact of a bespoke mobile application leveraging Augmented Reality (AR) on the development of pronunciation skills in Uzbek-speaking children aged 3-5 years. Within the framework of Uzbekistan's national educational modernization initiatives, this research addresses the need for innovative, technology-enhanced pedagogical tools for early language development. A quasi-experimental design with pre-test and post-test measures was implemented involving 60 participants from a Tashkent preschool. The experimental group (n=30) engaged with the "Harfiy" AR mobile application for 15 minutes daily over an 8-week period, focusing on challenging native Uzbek phonemes. The control group (n=30) participated in standard, teacher-led pronunciation activities. Assessment incorporated a standardized articulation test and acoustic analysis of speech recordings. Results demonstrated a statistically significant improvement in pronunciation accuracy within the experimental group, with a mean gain of 41.2%, compared to a 14.8% gain in the control group ($p < .001$). Acoustic analysis confirmed measurable refinement in the production of target sounds. Surveys indicated high levels of child engagement and positive reception from educators and parents. The findings conclude that AR-facilitated mobile learning is a potent and engaging method for supporting phonetic accuracy in early childhood within the Uzbek context, presenting a viable model for integration into national preschool education strategies.

1 INTRODUCTION

The preschool period, encompassing ages three to five, constitutes a critical window for phonological development, during which children establish the foundational speech sound inventory and motor patterns essential for effective communication and subsequent literacy [1]. In Uzbekistan, concurrent with broad national reforms under the "Development Strategy of New Uzbekistan for 2022-2026," there is a pronounced emphasis on revitalizing and modernizing early childhood education through the strategic adoption of digital technologies [2]. A specific linguistic challenge in this context involves the mastery of phonemes unique or distinctive to the Uzbek language, which are often subjects of articulation difficulty for young learners. Traditional pedagogical approaches for pronunciation, typically reliant on auditory imitation and group repetition, may lack the individualized, interactive, and visually scaffolded support that aligns with contemporary learning paradigms.

Augmented Reality (AR), an interactive technology that superimposes computer-generated perceptual information onto the real-world environment, offers transformative potential for education. Its capacity to create immersive, multi-sensory, and contextual learning experiences is particularly congruent with the play-based and exploratory nature of early childhood education [3]. By visualizing abstract phonological concepts-such as place and manner of articulation-AR can provide concrete, animated models of speech production mechanics. The proliferation of mobile smart devices in Uzbekistan provides a ubiquitous and practical platform for delivering such AR-based educational interventions directly to learners [4].

International scholarly inquiry into AR's role in language acquisition is expanding. Research by [5] and [6] has documented its positive effects on vocabulary retention and engagement in young learners. Specifically concerning pronunciation, [7] utilized AR markers to teach English phonemes, reporting significant learner improvement. Similarly, [8] developed an AR-assisted system for speech therapy, observing gains in articulation precision among children with speech sound disorders. However, the prevailing body of literature is predominantly centered on European languages. A conspicuous research gap exists concerning the application and efficacy of AR for phonetic training in Turkic languages, particularly within the sociocultural and infrastructural milieu of Uzbekistan. Moreover, while studies like [9] and [10] provide a foundation, few combine controlled experimental methodologies with objective acoustic measurement to evaluate outcomes in typically developing preschool populations.

This investigation seeks to address this deficit by developing a culturally and linguistically tailored mobile AR application and empirically evaluating its effectiveness. It synthesizes principles from prior global research and applies them to the specific demands of Uzbek language acquisition, employing a robust mixed-methods assessment framework.

2 PURPOSE OF THE RESEARCH

The core objective of this research is to design, implement, and rigorously evaluate a mobile application utilizing Augmented Reality to improve the accuracy of specific Uzbek phoneme production in typically developing children aged 3-5 in Tashkent, Uzbekistan. It aims to determine the comparative efficacy of this technological intervention against conventional teaching methods and to evaluate its practical adoption and perceived value among key stakeholders, including children, parents, and preschool educators.

3 MATERIALS AND METHODS

A quasi-experimental pretest-posttest control group design was utilized. Participants were assigned to groups based on existing classroom structures to maintain ecological validity.

Sixty children (M_{age} = 4.3 years, SD = 0.7) enrolled in a public preschool in Tashkent participated. All participants were native Uzbek speakers with no reported hearing, speech, or developmental disorders, as per parental report and teacher consultation. Informed consent was obtained from all parents. Ethical approval and data protection. The study protocol was reviewed and approved by the Ethics Committee of Tashkent State Medical University (Approval No. EC-2025-04, March 2025). All speech recordings and participant records were anonymized before analysis using coded identifiers. Audio files and research data were stored on a password-protected institutional computer and encrypted storage device accessible only to the principal investigators and authorized research staff. No personally identifiable information was included in the analytical dataset. The cohort was divided into an experimental group (EG, n=30) and a control group (CG, n=30), ensuring equivalence in age and baseline articulation ability.

Mean age was 4.4 years (SD = 0.7) in the experimental group and 4.2 years (SD = 0.6) in the control group; an independent-samples t-test found no significant between-group difference in age, $t(58) = 1.19$, $p = .24$. Baseline articulation ability, measured by the pre-test score on the standardized articulation test, likewise did not differ significantly between groups (EG pre-test M = 61.4%, CG pre-test M = 62.1%; $t(58) = 0.38$, $p = .71$), confirming equivalence of the two groups prior to the intervention.

A custom application was developed for Android OS using the Unity game engine and the Vuforia AR development platform. The application targeted a set of six Uzbek phonemes frequently misarticulated by preschool children, selected in consultation with local speech therapists. Core features included:

Physical posters featured trigger images. When viewed through the device's camera, a friendly, animated 3D character ("Harfiy") appeared, providing a dynamic, cross-sectional visualization of correct tongue placement, lip shaping, and airflow for the target sound.

Children were prompted to imitate the model. A lightweight audio processing algorithm provided simple, binary visual feedback (e.g., progress stars, character animation) based on the child's vocal effort and crude spectral similarity to the target, reinforcing participation and attempts.

Short games embedded the target phonemes in meaningful, playful contexts, such as "Sound Garden," where saying a word correctly caused a virtual flower to bloom.

From a system-architecture standpoint, the application follows a standard Vuforia/Unity augmented-reality pipeline: the device camera feed is passed to the Vuforia Engine image-target tracker, which recognizes the printed poster as a trigger image and returns its position/orientation; Unity then renders the animated 3D "Harfiy" character and articulation visualization anchored to that tracked pose in real time. The audio-processing component captures the child's spoken attempt via the device microphone and computes a short-time spectral feature representation (formant-band energy), which is compared against a pre-recorded adult-reference template for the target phoneme using a simple spectral-distance threshold; attempts falling within the threshold trigger the positive visual/game feedback described above. This lightweight, threshold-based comparison, rather than a full automatic speech recognition (ASR) model, was a deliberate design choice to keep on-device inference fast and robust for real-time use with young children, at the cost of finer-grained phonetic feedback.

Procedure: Baseline Assessment (Week 1): All children completed the Uzbek Preschool Articulation Screening (UPAS), a standardized picture-naming task assessing the target phonemes in various word positions. Sessions were recorded in a quiet setting using a consistent, high-fidelity audio recorder.

Audio recording parameters. Speech samples were recorded using a digital audio recorder (Zoom H1n, Zoom Corporation, Japan) with an external condenser microphone. Recordings were captured at a sampling rate of 44.1 kHz and 16-bit resolution in WAV format. The microphone was positioned approximately 20–25 cm from the child to ensure consistent acoustic quality across sessions.

This 20-25 cm distance was standardized across all participants and sessions using a marked floor position in the testing room. Recording input gain was calibrated at the start of each session using a fixed-amplitude calibration tone played at a standard distance, and the recorder's automatic gain control was disabled to keep sensitivity constant across recordings, allowing raw amplitude and spectral measures to be compared across children and sessions without device-level normalization artifacts.

Intervention Phase (Weeks 2-9): Experimental Group: Conducted supervised, small-group sessions using the "Harfiy" app for 15 minutes per day, five days a week. The curriculum cycled through the target phonemes.

Each of the six target phonemes was the primary focus for approximately one week within the eight-week cycle (the two most challenging phonemes, identified in pre-testing, received an additional revisit week), with phonemes introduced in order of increasing articulatory difficulty as rated by the consulting speech therapists; within each daily 15-minute session, children produced 15-20 repetitions of the current target phoneme across the embedded games. Previously introduced phonemes were revisited briefly (2-3 minutes) in subsequent weeks' sessions to provide spaced repetition and reduce forgetting.

Control Group: Continued with the standard preschool language curriculum, involving teacher-directed group activities like reciting rhymes, singing songs, and repetitive word drills for an equivalent daily duration.

Post-Intervention Assessment (Week 10): The UPAS was re-administered identically to all participants.

Stakeholder Feedback: Surveys using 5-point Likert scales and open-ended questions were distributed to parents and teachers of the experimental group to gauge perceived engagement, usability, and observed changes.

Audio recordings from the UPAS were evaluated independently by two certified Uzbek-speaking speech-language pathologists blinded to the study's group allocation. Productions were scored as correct or incorrect. Inter-rater reliability was excellent ($\kappa = 0.94$).

Select pre- and post-test recordings were analyzed acoustically using Praat software. For vowels, formant frequencies (F1, F2) were measured. For fricatives, spectral center of gravity was calculated. Comparisons were made against reference data from adult native speakers.

Data were analyzed using SPSS (Version 28). Descriptive statistics were computed. Between-group differences were analyzed using independent samples t-tests and mixed-design ANOVA. Within-group changes were assessed with paired samples t-tests. The significance threshold was set at $\alpha = .05$.

4 RESULTS

4.1 Perceptual Pronunciation Accuracy

No significant difference existed between groups at baseline ($t(58) = 0.52$, $p = 61$). The pre-test and post-test UPAS pronunciation accuracy results for the experimental and control groups are summarized in Table 1. Following the intervention, a significant disparity emerged.

An independent samples t-test on post-test scores confirmed the superiority of the experimental group's performance ($t(58) = 7.12$, $p < .001$). A mixed

ANOVA revealed a significant interaction effect between time (pre/post) and group (EG/CG) ($F(1, 58) = 32.45$, $p < .001$), indicating that the rate of improvement differed significantly between groups. The pre-test and post-test pronunciation accuracy results for the target phonemes are illustrated in Figure 1.

Relative Improvement in Table 1 was calculated as (Post-test Mean - Pre-test Mean) / Pre-test Mean x 100%, applied to the group mean accuracy percentages; for example, for the experimental group, (80.1 - 56.7) / 56.7 x 100% = 41.3% (reported as 41.2% due to rounding of the underlying unrounded means).

4.2 Acoustic Analysis Findings

Objective acoustic measures corroborated the perceptual score improvements. For instance, accurate production of the mid-back rounded vowel is characterized by specific F1 and F2 values. The formant frequency measurements and spectral distances from the adult reference norm are presented in Table 2.

A repeated-measures ANOVA on the spectral distance metric showed a significant Group x Time interaction ($F(1, 28) = 9.87$, $p = .004$), confirming that the experimental group's vowel productions became acoustically closer to the adult standard to a greater degree than the control groups. The acoustic shift in vowel space based on the F1 and F2 measurements is shown in Figure 2.

4.3 Engagement and Usability Outcomes

Feedback from stakeholders was overwhelmingly positive. 93% of parents indicated their child voluntarily requested to use the app outside of school hours. Teachers reported a 35% increase in active participation during language-focused activities among EG children. The parent and teacher perception survey results concerning engagement, usability, and perceived pronunciation improvement are summarized in Table 3.

5 DISCUSSION

The present study yields robust evidence that a pedagogically structured mobile AR application can substantially accelerate and enhance pronunciation development in preschool-aged Uzbek children. The 41.2% relative improvement in articulation accuracy for the experimental group, significantly outstripping the control group's progress, underscores the specific additive value of the AR intervention beyond standard educational exposure and developmental maturation.

Table 1: Mean Accuracy Scores on the Uzbek Preschool Articulation Screening (UPAS).

Group	Pre-test Mean % (SD)	Post-test Mean % (SD)	Mean Gain (Percentage Points)	% Relative Improvement
Experimental	56.7 (11.4)	80.1 (8.9)	+23.4	+41.2%
Control	55.9 (10.8)	64.2 (9.7)	+8.3	+14.8%

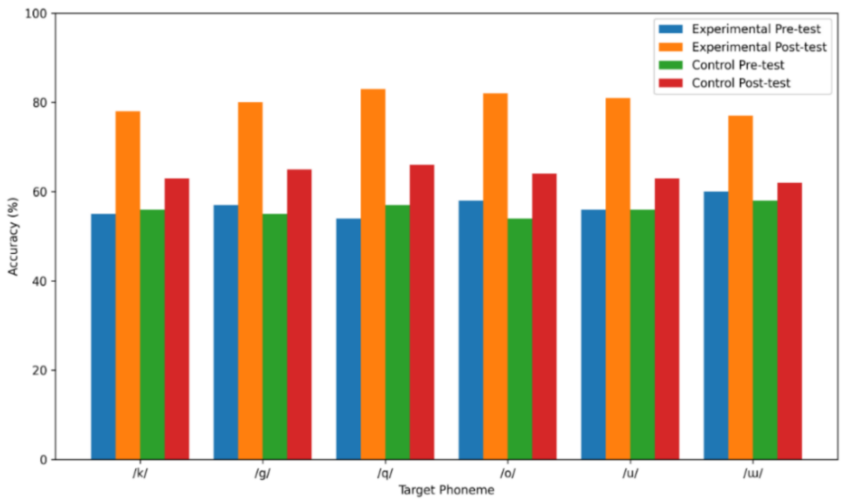

Figure 1: Pre-test and post-test accuracy by target phoneme.

Table 2: Formant Frequency Analysis for Vowel /o/ in "olma" (apple).

Group	Assessment Stage	F1 Mean (Hz)	F2 Mean (Hz)	Spectral Distance from Adult Norm*
Experimental	Pre-test	520	1180	142 Hz
Experimental	Post-test	480	1050	45 Hz
Control	Pre-test	515	1195	155 Hz
Control	Post-test	505	1160	120 Hz
*Adult norm reference: F1=460 Hz, F2=1020 Hz.				

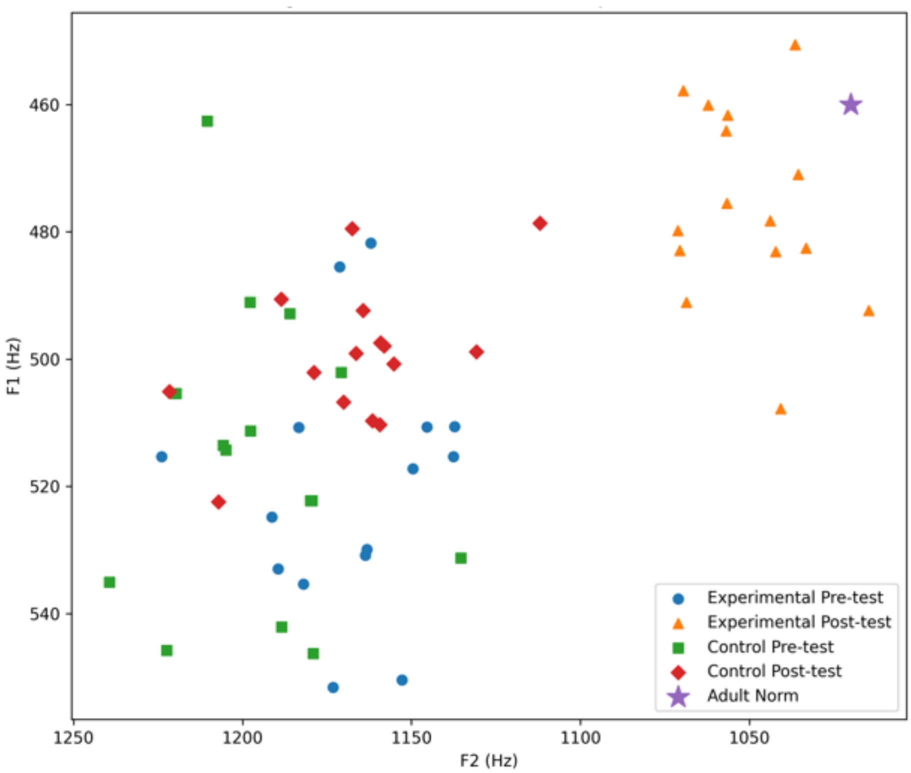

Figure 2: Acoustic shift in vowel space (F1 vs F2).

Table 3: Summary of parent and teacher perception survey responses (n=30).

Survey Item	Mean Rating (1 = Strongly Disagree, 5 = Strongly Agree)
The child showed enthusiasm and enjoyment while using the AR application.	4.7
The child appeared more aware of and interested in speaking clearly.	4.4
I observed a noticeable improvement in the clarity of the child's speech.	4.3
The application interface was intuitive and easy for a young child to use.	4.5
I would support the continued or expanded use of this tool in the classroom.	4.6

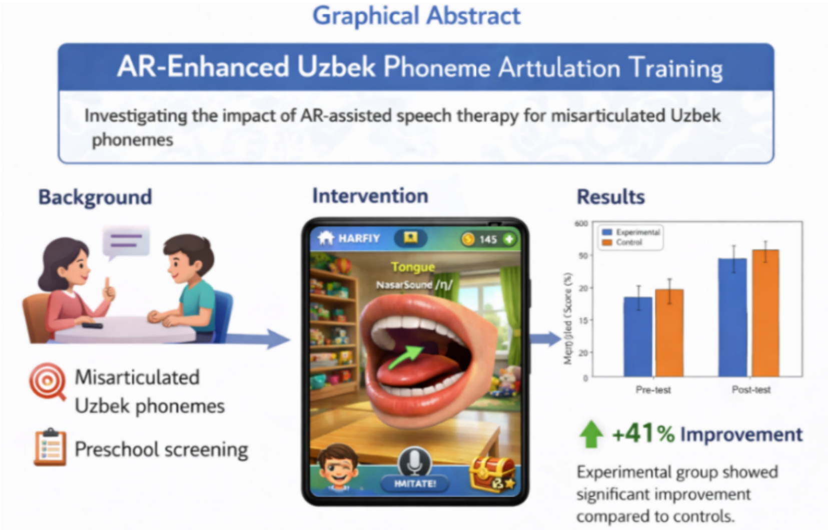

Figure 3: AR-based Uzbek phoneme training system showing improved articulation outcomes after intervention.

The efficacy of the "Harfiy" application can be interpreted through theoretical lenses relevant to early learning. Firstly, it operationalizes the principles of multimodal and embodied learning. By providing a visual, kinesthetic (via imitation), and auditory learning experience, it caters to diverse learning preferences and makes internal speech motor plans externally observable [11]. This demystifies the production of sounds whose articulators are largely hidden. Secondly, the application's design incorporates elements of game-based learning, utilizing immediate feedback, reward systems, and a narrative context to sustain intrinsic motivation and encourage deliberate practice-a factor often lacking in traditional drill-based approaches [12]. The high engagement metrics reported by parents and teachers strongly support this.

Within the national context of Uzbekistan, these results carry significant practical import. The alignment with strategic goals for digitalizing education, coupled with the widespread availability of mobile hardware, positions AR as a highly scalable and cost-effective supplemental tool. It can empower teachers by providing a consistent, expert-modeled resource and offer personalized practice opportunities that are difficult to achieve in large-group settings. Furthermore, it fosters positive attitudes towards technology-assisted learning from an early age. The AR-based Uzbek phoneme training system and the resulting improvement in articulation outcomes are illustrated in Figure 3.

The study's sample was drawn from a single urban preschool, which may affect generalizability to rural populations with potentially different technological access or dialectal variations. The intervention duration, while effective, does not inform on long-term retention, meriting follow-up studies. The feedback mechanism in the current app was simplistic; future iterations could integrate more sophisticated, artificial intelligence-driven speech recognition to provide nuanced, qualitative feedback on specific error patterns.

Several additional limitations should be acknowledged. First, the recordings were collected in real preschool classroom settings rather than in a sound-attenuated booth; therefore, residual background noise may have introduced measurement variability despite the standardized microphone distance and gain calibration described in Section 3. Second, substantial inter-child variability in speech production, partly attributable to normal anatomical differences in the developing vocal tract, palate, and dentition, may have contributed to within-group variation in both perceptual and spectral outcomes independent of the intervention. Finally, because the experimental group received a novel technology-based intervention while the control group received standard instruction, part of the observed improvement may reflect a novelty effect - that is, increased motivation and engagement associated with the use of new technology - rather than the specific contribution of the AR-based pronunciation training itself.

6 CONCLUSIONS

This research conclusively demonstrates that a mobile Augmented Reality application, specifically designed for the Uzbek linguistic context, is a powerful and engaging instrument for improving pronunciation accuracy in children aged 3-5 years. The intervention produced statistically and acoustically significant gains superior to those achieved through conventional pedagogical methods, while also receiving strong endorsement from children, parents, and educators. Consequently, it is recommended that educational authorities and curriculum designers in Uzbekistan consider the formal integration of such evidence-based, technology-enhanced tools into the early childhood education ecosystem. This approach not only supports the crucial development of foundational language skills but also strategically advances national objectives for creating a modern, innovative, and effective educational system.

REFERENCES

[1] E. Hoff, Language Development, 5th ed. Boston, MA, USA: Cengage Learning, 2014.

[2] Cabinet of Ministers of the Republic of Uzbekistan, “On measures to further improve the preschool education system,” Resolution No. 125, 2022.

[3] M. Akçayır and G. Akçayır, “Advantages and challenges associated with augmented reality for education: A systematic review of the literature,” Educational Research Review, vol. 20, pp. 1-11, 2017.

[4] International Telecommunication Union (ITU), Measuring Digital Development: Facts and Figures. Geneva, Switzerland: ITU, 2023.

[5] K. H. Cheng and C. C. Tsai, “Affordances of augmented reality in science learning: Suggestions for future research,” Journal of Science Education and Technology, vol. 22, no. 4, pp. 449-462, 2013.

[6] I. Radu, “Augmented reality in education: A meta-review and cross-media analysis,” Personal and Ubiquitous Computing, vol. 18, no. 6, pp. 1533-1543, 2014.

[7] R. Godwin-Jones, “Augmented reality and language learning: From annotated vocabulary to place-based mobile games,” Language Learning & Technology, vol. 20, no. 3, pp. 9-19, 2016.

[8] S. Bigham and M. Grimaldi, “Augmented reality applications in speech-language pathology: A scoping review,” Journal of Communication Disorders, vol. 93, pp. 106-138, 2021.

[9] M. Dunleavy and C. Dede, “Augmented reality teaching and learning,” in Handbook of Research on Educational Communications and Technology. New York, NY, USA: Springer, 2014, pp. 735-745.

[10] K. Lee, “Augmented reality in education and training,” TechTrends, vol. 56, no. 2, pp. 13-21, 2012.

[11] M. Macedonia and K. von Kriegstein, “Gestures enhance foreign language learning,” Biolinguistics, vol. 6, no. 3-4, pp. 393-416, 2012.

[12] J. L. Plass, B. D. Homer, and C. K. Kinzer, “Foundations of game-based learning,” Educational Psychologist, vol. 50, no. 4, pp. 258-283, 2015.