VoiceBridge: An Adaptive Bilingual Speech Therapy Support Application Using Context-Aware Deep Neural Networks

Abstract

Speech impairments significantly affect communication, educational participation, and social inclusion, particularly in low-resource settings where access to professional speech therapists is limited. This work presents VoiceBridge, a bilingual mobile application designed to assist individuals with speech difficulties through adaptive, deep neural network techniques. The system supports English and Twi languages and follows structured speech therapy methodologies, beginning with phonemic awareness using alphabets and numbers and progressively advancing to syllables and contextual speech exercises. The application processes user speech input using acoustic feature extraction techniques and evaluates pronunciation accuracy through a deep neural network-based speech classification model trained on reference phoneme patterns.

Keywords: Assistive Artificial Intelligence, Speech Recognition, Deep Neural Networks, Adaptive Learning Systems, Bilingual Speech Technology

Introduction

Speech and communication disorders create barriers to effective interaction, academic performance, and social inclusion (Trikoilis and Billiri 2024). In many communities, particularly in resource-constrained environments, access to trained speech therapists is limited due to cost, availability, and geographic constraints (Botes, M. 2026). Many of these digital learning tools lack personalization and do not support local languages, limiting their effectiveness in multilingual contexts(Irewole et al. 2025). This research introduces VoiceBridge, an adaptive bilingual speech therapy support application that leverages deep neural networks to personalize speech training. The objective is to create a system that mirrors structured therapeutic progression while dynamically adjusting to individual user performance. By integrating English and Twi languages, the project promotes linguistic inclusivity and expands equitable access to speech support technologies.

Methods

VoiceBridge follows a modular architecture comprising four primary components: a user interaction interface, a speech-processing module, a deep neural network-based pronunciation evaluation engine, and an adaptive learning mechanism. The mobile interface provides guided exercises and visual feedback that help users practice phoneme articulation through structured therapy sessions. User speech recordings are first processed through an acoustic feature extraction pipeline. The system uses Mel-Frequency Cepstral Coefficients (MFCCs), along with spectral and temporal features, to represent phonetic information in a compact numerical form suitable for machine learning models.

These features capture vocal tract characteristics, frequency variations, and phoneme articulation patterns, which are essential for accurate pronunciation evaluation. The extracted features are passed into a deep neural network classifier trained on reference phoneme recordings. The network learns complex representations of speech signals and predicts pronunciation similarity between user input and reference speech patterns. The output layer generates a pronunciation accuracy score that reflects how closely the spoken phoneme matches the expected articulation. An adaptive learning engine analyzes user performance metrics, including pronunciation accuracy, response time, and repetition consistency. Based on these metrics, the system dynamically adjusts exercise difficulty, repetition frequency, and phonetic focus areas to reinforce weaker articulation patterns while gradually introducing more complex speech exercises.

Results

Preliminary prototype testing was conducted using multiple speech practice sessions to evaluate system performance under adaptive learning conditions. The evaluation focused on pronunciation accuracy, response time, and exercise completion rate across successive training sessions.

Parameter	**Session 1**	**Session 2**	**Session 3**
Pronunciation Accuracy	0.80	0.87	0.93
Average Response Time (seconds)	13.8	11.2	9.6
Exercise Completion Rate	90%	94%	97%

The results indicate that adaptive training improves both speech articulation accuracy and user engagement over repeated sessions. Pronunciation accuracy increased steadily across sessions while response time decreased, suggesting improved familiarity and confidence during speech exercises.

Discussion and Conclusion

This study introduced VoiceBridge, an adaptive bilingual speech therapy support application that integrates deep neural networks and context-aware learning mechanisms to assist individuals with speech articulation difficulties. By supporting both English and Twi, the system addresses a critical gap in speech therapy technologies for multilingual and under-resourced communities. Preliminary results demonstrate that adaptive learning strategies can improve pronunciation accuracy, reduce response time, and enhance user engagement during speech practice sessions.

References

Trikoilis, Dionysios, and Kalliopi Billiri. 2024. “Social and Language Development Interventions Regarding Adolescents With Autism Spectrum Disorder.” *Turkish Journal of Special Education Research and Practice*, April. https://doi.org/10.37233/trsped.2024.0147.

Botes, M. (2026). Enhancing language development in low-income communities through pre-recorded therapy videos: equipping speech-language therapy students with digital intervention skills. *European Early Childhood Education Research Journal*, 1–14. https://doi.org/10.1080/1350293X.2026.2619578

Irewole, Michael, Abegunde Kolawole, Arogundade Oluwaseun, and Festus Solanke. 2025. “Multilingual Learner in Digital Age: Investigating the Effectiveness of E-learning in Language Acquisition and Academic Achievement.” *International Journal of English Teaching and Learning.* 3 (4): 82–92. https://doi.org/10.11648/j.ijetl.20250304.12.