International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Comprehensive Review Paper on SpeechEase: AI-Based Multilingual Speech Therapy System

[1]Prof. Monika S. Shirbhate, [2]Dr. Rani M. Khandare, [3]Dip Rajesh Kuralkar,
[4]Nandini Kiran Yeul, [5]Mohmmad Faijan Rashid Sheikh, [6]Shubham Tanaji Yedle
[1,2]Assistant Professor, [3,4,5,6]Bachelor of Engineering Scholars Department
of Information Technology
Prof. Ram Meghe College of Engineering and Management Badnera-Amravati India Sant
Gadge Baba Amravati University

***Abstract*—Speech and pronunciation difficulties can affect communication, learning, academic participation, and social interaction. Conventional speech assessment and therapy generally depend on trained professionals and repeated practice sessions, which may not always be easily accessible to all users. Recent developments in Artificial Intelligence (AI), Automatic Speech Recognition (ASR), machine learning, deep learning, and speech signal processing have created opportunities for automated speech assessment and personalized pronunciation practice. This paper reviews AI-assisted speech technologies with emphasis on multilingual speech processing, speech-sound analysis, personalized feedback, and progress monitoring. It follows a focused semi-systematic/narrative review approach and discusses selected literature on automatic speech analysis, personalized pronunciation correction, and articulatory analysis. The reviewed work indicates that AI may support speech practice outside continuous clinician supervision; however, language coverage, phoneme-specific evaluation, dataset availability, and integration of assessment with longitudinal practice remain challenges. The review identifies a research gap in combining multilingual pronunciation assessment, weak-sound or word identification, targeted practice, repeated re-assessment, and day-wise progress tracking in one platform. An initial proposed framework is outlined as a future development direction. No application has been implemented or experimentally evaluated at this stage. The proposed system is intended to support practice and early intervention, not to replace professional speech-language assessment or clinical therapy.**

***Index Terms*—Artificial Intelligence, Speech Therapy, Multilingual Speech Processing, Automatic Speech Recognition, Pronunciation Assessment, Machine Learning, Deep Learning, Speech Signal Processing, Personalized Feedback**

I. Introduction

Speech is one of the most important methods of human communication. Accurate production of speech sounds is necessary for intelligibility and effective interaction. Difficulties in producing particular speech sounds may lead to substitutions, omissions, distortions, or other pronunciation deviations. Such difficulties can affect communication, learning, confidence, and participation in everyday activities. Traditional speech assessment and therapy are generally conducted by qualified speech-language professionals. These approaches can provide individualized assessment and intervention; however, they may require regular appointments, professional availability, and continuous practice outside clinical sessions.

Technology-based solutions are therefore being explored to provide additional opportunities for speech practice and monitoring. Artificial Intelligence has become an important technology in speech and language applications, with machine-learning and deep-learning approaches being studied for speech classification, analysis, and enhancement across multiple speech and language disorders [1]. AI systems can process speech signals, recognize spoken content, identify acoustic characteristics, and perform classification or prediction tasks. ASR systems convert speech into machine-readable representations, while machine-learning and deep-learning models can be used to analyze pronunciation and speech-sound characteristics.

Recent studies demonstrate the potential of automated speech analysis in speech-sound-disorder applications. Carl et al. evaluated an automatic speech-sound analysis algorithm against clinician assessment using 3,584 consonant stimuli produced by 395 American English-speaking children and adolescents with speech sound disorders. Their findings showed high agreement for correctly produced sounds, while performance for misarticulated sounds and variation across phoneme complexity remained important limitations [2].

AI-assisted intervention has also been investigated. Benway and Preston described an AI-assisted /r/ therapy intervention in a single-case experimental study involving ten 40-min sessions, showing the feasibility of integrating automated perceptual estimation into motor-based intervention [3]. Lakshminarayanan et al. proposed personalized pronunciation correction using a customized Proximal Policy Optimization method together with a large language model to generate individualized feedback [4]. For multilingual recognition, Tripathi et al. proposed language-family prompt tuning and a customized tokenizer for Whisper, targeting Indian languages and addressing both recognition accuracy and inference speed [5]. Articulatory analysis has also progressed: Benway et al. used speech-inversion neural networks to relate inferred vocal-tract kinematics to perceptual ratings of /r/ and /s/ in childhood speech sound disorders [6].

Although these studies demonstrate progress, the research is distributed across different tasks: assessment, correction, recognition, and articulatory analysis. An integrated multilingual platform could connect initial assessment, identification of sounds or words requiring practice, targeted exercises, reassessment, and progress history. This paper reviews relevant

approaches and identifies gaps that can inform the planning of such a system.

The work is at the initial stage; the proposed framework is a plan, not a completed implementation.

II. Review Methodology

A. Review Approach

A literature review is a research methodology for identifying, evaluating, and synthesizing existing knowledge rather than simply listing papers. Snyder distinguishes systematic, semi-systematic, and integrative review approaches and discusses their appropriate uses and methodological requirements [7]. A systematic review uses predefined and reproducible procedures; a semi-systematic review is useful for broad, interdisciplinary topics with varying terminology; and an integrative review combines different types of evidence to build a wider conceptual understanding.

Because AI-assisted speech support spans speech-language pathology, artificial intelligence, machine learning, speech processing, and human-computer interaction, this paper is framed as a focused semi-systematic/narrative review. It is intended to establish background and research direction at the project-planning stage, not to claim an exhaustive systematic search.

B. Review Questions

The primary question is: What AI and speech-processing approaches have been used for automated pronunciation or speech-sound assessment and personalized speech support, and what gaps remain for an integrated multilingual speech-practice platform? Supporting questions consider the AI techniques used, automatic identification of speech-sound errors, generation of personalized feedback, languages and datasets represented, limitations of current systems, and possible integration of assessment, practice, reassessment, and progress tracking.

C. Literature Focus and Data Extraction

The review focuses on scholarly work concerning AI in speech and language applications, ASR, automatic speech analysis, pronunciation assessment and correction, speech sound disorders, rehabilitation, personalized practice, multilingual processing, and articulatory analysis. Recent studies, particularly those published from 2024 to 2026, are emphasized, while earlier methodological literature is included where relevant.

For each study, the review considers its objective, method, participants or dataset, language, contribution, limitations, and relevance to the proposed research direction.

The source material available for this draft provides selected studies and bibliographic details. Before formal submission, the project team should verify each reference against the publisher or digital library record and add the exact databases, search strings, search dates, and inclusion/exclusion criteria if the college requires a reproducible review protocol.

III. Background of AI-Assisted Speech Support

A. Artificial Intelligence in Speech Applications

Artificial Intelligence refers to computational methods that enable machines to perform tasks involving learning, recognition, prediction, classification, and decision-making. In speech applications, AI can analyze audio signals and learn relationships between speech characteristics and expected outputs. Machine-learning models may classify speech patterns, while deep-learning models can learn complex representations from speech datasets. A broad 2025 review of AI for speech and language disorders synthesized evidence across classification and enhancement tasks and illustrates the breadth of current AI-based speech research [1].

B. Automatic Speech Recognition

Automatic Speech Recognition converts human speech into a machine-readable representation, commonly text or another linguistic representation. ASR can provide useful information about what a speaker has said. For pronunciation assessment, however, correct transcription alone may not be sufficient: a word may be recognized correctly even when its production contains a phonetic or articulatory deviation. Ahn et al. specifically note that general-purpose ASR systems trained to predict standard spelling are not necessarily suitable for diagnosing pronunciation in children with speech sound disorders [8].

For multilingual applications, Tripathi et al. showed that prompt tuning with language-family information and customized tokenization can improve Whisper performance for Indian languages, demonstrating the importance of language-specific adaptation [5]. Kim et al. similarly evaluated an ASR system based on a fine-tuned wav2vec 2.0 XLS-R model for children with speech sound disorders and reported that performance varies according to phoneme or word characteristics [9].

C. Speech Signal Processing

Speech signal processing examines acoustic properties of recorded speech. Relevant characteristics can include frequency, pitch, energy, duration, spectral features, formant information, and temporal patterns. Preprocessing may include noise reduction, silence removal, normalization, segmentation, and sampling-rate conversion. The choice of processing steps depends on the recording conditions and the model being used. Such processing may help make recordings more consistent, but it cannot by itself guarantee accurate assessment.

D. Phoneme-Level Analysis

Pronunciation is closely related to the production of phonemes—the sound units that distinguish words in a language. A phoneme-level approach compares a user's production with an expected target sound. It may help identify sounds that require additional practice. Evidence from Hindi pronunciation research is especially relevant here: Rustagi et al. proposed Dhvani, a weakly supervised phonemic error-detection and personalized-feedback system for Hindi, with analysis

focused on phonemic distinctions rather than only word transcription [10]. Self-supervised wav2vec2-based speech-rating work has also explored automatic assessment for children with speech sound disorder [11].

The comparison must account for language-specific phoneme inventories and pronunciation variation; an English-oriented model cannot automatically be assumed to work equally well for Hindi or other Indian languages. For text resources used in multilingual language technology, the AI4Bharat-IndicNLP Corpus provides monolingual corpora and word embeddings for Indic languages, but such text resources should not be treated as substitutes for labeled child speech corpora [12].

E. Deep Learning, LLMs, and Reinforcement Learning

Deep-learning models can learn complex patterns from speech data and have been used for recognition, classification, pronunciation assessment, and articulatory analysis. Their performance may be affected by dataset size, speaker diversity, language characteristics, recording conditions, and annotation quality. LLMs can generate natural-language feedback, while reinforcement learning can optimize decisions based on feedback or performance. Lakshminarayanan et al. combine reinforcement learning with an LLM for personalized pronunciation correction [4]. Such methods are relevant as possible components, but their suitability for a child-focused multilingual application would require evaluation.

IV. Review of Selected Existing Studies

A. Automatic Speech Analysis

Carl et al. compared an automatic speech-sound analysis algorithm with clinician assessment. Their study included 3,584 consonant stimuli produced by 395 American English-speaking children and adolescents with speech sound disorders. The study reported agreement above 80

A complementary validation study by Kim et al. compared ASR-based assessment with clinician transcriptions for Korean children with speech sound disorders. The model was based on wav2vec 2.0 XLS-R and was additionally trained using child speech and articulation-error data. The authors reported potential for ASR-based assessment, while also noting variability across phonemes and word characteristics and the need for broader speech data for clinical applicability [9].

B. Personalized Pronunciation Correction

Lakshminarayanan et al. proposed an automated pronunciation-correction system using reinforcement learning and LLMs. The approach uses a customized Proximal Policy Optimization method for pronunciation evaluation and an LLM to provide individualized feedback [4]. The contribution is its emphasis on feedback tailored to the user rather than only a correct/incorrect label. For the present project direction, the work raises questions about how personalized feedback can be linked to identified weak sounds and structured exercises. Multilingual adaptation and suitability for children would need to be established through further evidence.

C. Articulatory Speech Analysis

Benway et al. investigated whether articulatory characteristics inferred by speech-inversion neural networks correspond to perceptual ratings of /r/ and /s/ in children with speech sound disorders. The study analyzed 5,961 utterances from 118 children and three adults. The findings indicated that inferred vocal-tract variables differentiated categories and magnitudes of articulatory errors, particularly for /r/ [6]. This indicates that AI-based speech analysis may provide information beyond simple recognition. The study addresses selected sounds and does not itself provide a complete multilingual platform with practice, reassessment, and day-wise tracking.

D. Broader AI Applications in Speech-Language Pathology

A 2024 review by Bhardwaj et al. discusses the evolving role of AI, machine learning, and augmentative and alternative communication in pediatric speech and language disorders, including diagnosis and therapy applications [13]. A systematic review by Shi et al. examined 21 commercially available AI-enabled mobile apps for pediatric speech therapy and evaluated them using 15 criteria, highlighting opportunities as well as ethical and design concerns [14]. Dadgar et al. provided a broader scoping review of AI-driven technologies for pediatric speech and language therapy and emphasized the role of ASR and audio-processing technologies while identifying implementation and evidence gaps [15].

The selected studies cover distinct aspects of the problem: automated evaluation, personalized feedback, articulatory analysis, multilingual ASR, digital practice, and broader AI applications. They provide background for a proposed system, but they should not be interpreted as proof that the planned application already exists or has achieved measurable results.

V. Comparative Analysis of Selected Research

Table I summarizes the selected studies based on the information available in the source material. The limitations are described in relation to the proposed project scope and should not be read as a complete critique of each original publication. The comparison suggests that current research can inform individual components, but the proposed project would need to connect these components and evaluate them together. Multilingual ASR may help recognize spoken content, but a separate pronunciation-analysis method would be needed to assess whether target sounds were produced as intended. Similarly, personalized feedback is useful only if it is appropriate, understandable, and linked to practice tasks that can be evaluated over time.

VI. Discussion

The reviewed studies demonstrate that AI can contribute to several stages of speech assessment and practice. Automatic speech analysis can provide judgments that may be compared with clinician assessment [2], [9]. Reinforcement learning and LLMs offer one possible route to individualized pronunciation feedback [4]. Speech-inversion neural networks may provide

TABLE I
COMPARATIVE ANALYSIS OF SELECTED RESEARCH

Study	Focus/Contribution	Limitation	Relevant to This Review
Benway et al. (2026) [6]	Speech-inversion neural networks and perceptual ratings used to analyze articulatory characteristics of /r/ and /s/ errors in children.	Selected sounds and analysis task; not a complete multilingual assessment-to-practice platform.	Articulatory analysis
Remya et al. (2025) [1]	Review of AI techniques for speech classification and enhancement across multiple speech and language disorders.	Broad scope; not focused on a single multilingual child-practice workflow.	AI background
Tripathi, Gothi, and Wasnik (2025) [5]	Whisper-based multilingual ASR for Indian languages using language-family prompt tuning and a customized tokenizer.	Recognition improvement is not the same as mispronunciation assessment or personalized therapy exercise generation.	Indian-language ASR
Bhardwaj et al. (2024) [13]	Review of AI, ML, and AAC technologies in pediatric speech and language disorders; discusses diagnosis, therapy, and early intervention.	Identifies opportunities but does not implement the proposed integrated application.	Pediatric AI context
Carl et al. (2024) [2]	Compared automatic speech-sound analysis with clinician assessment and investigated potential use for speech-sound practice.	Primarily focuses on automatic analysis; not the full multilingual personalized practice and progress pipeline.	Automated assessment
Benway and Preston (2024) [3]	AI-assisted speech therapy research focused on /r/ production and a single-case experimental design.	Focused on one English sound and a limited participant sample; not a multilingual platform.	AI-assisted therapy
Deka et al. (2025) [16]	Systematic review of AI-based automated speech therapy tools, intervention modes, autonomy levels, and effectiveness evidence.	Evidence remains heterogeneous and few studies directly compare automated tools with conventional therapy.	System-level evidence
Hong et al. (2025) [18]	MORA uses story-based interactive practice, visual cues, feedback, and an AI-in-the-loop workflow involving speech-language pathologists.	Preprint; proposed system and formative/expert evaluation rather than a longitudinal clinical efficacy trial.	Clinic-to-home practice

articulatory information that is not available from word transcription alone [6]. Multilingual ASR and phonemic error-detection research further indicate the importance of language-specific modeling for Indian-language applications [5], [10].

At the same time, the literature is fragmented across different research questions and populations. A system that recognizes words is not necessarily a system that detects misarticulation. A model evaluated on selected English phonemes cannot be assumed to work for Hindi or Marathi without suitable data and testing. A feedback-generation method also does not automatically provide an age-appropriate, safe, and effective practice plan. Reviews of pediatric AI and automated therapy repeatedly identify evidence, usability, stakeholder involvement, and ethical concerns that should be considered during system design [13]–[16].

Digital practice systems also show why longitudinal evaluation matters. McLeod et al. evaluated the SayBananas! mobile game for rural Australian children and reported high ratings for engagement, functionality, aesthetics, and quality, while app use was associated with speech-production improvement over a four-week period [19]. Hair et al. evaluated Apraxia World over multiple treatment blocks with ten children and reported therapeutically significant speech improvements, increased engagement, and useful automated pronunciation evaluation within the game [20]. These findings motivate the inclusion of repeated practice and progress records, while the present project should still avoid assuming that similar outcomes will occur without its own evaluation.

For the planned application, a complete workflow could begin with a user selecting a language and reading or speaking prompted sounds and words. The system could analyze recordings, identify possible areas for practice, provide exercises, and invite the user to repeat the assessment. Storing dated results could help users and caregivers review practice history. Interactive clinic-to-home practice concepts such as MORA further illustrate how targeted content, feedback, and clinician involvement can be connected within one workflow [18]. These are proposed functions only; their technical feasibility and effectiveness remain to be tested.

The system should communicate that its output is an AI-assisted practice indication rather than a clinical diagnosis. Speech development varies, and recordings may be affected by accent, age, microphone quality, background noise, and other factors. Professional assessment remains important for clinical decisions.

VII. RESEARCH GAPS

A. Multilingual Speech Assessment

A substantial portion of the selected pronunciation research focuses on English. Languages differ in phoneme inventories, phonotactics, pronunciation patterns, accents, and available datasets. Models developed for English cannot automatically be considered suitable for Indian languages. Evidence from Whisper adaptation for Indian languages, Hindi phonemic error detection, and child-speech ASR demonstrates that language-specific and population-specific evaluation is necessary [5], [8]–[10]. Hindi and English are proposed as initial

target languages for the planned system, with additional languages considered only after the initial approach is evaluated.

B. Integration of System Components

Existing work may focus on speech recognition, speech classification, pronunciation assessment, articulatory analysis, or feedback generation as separate tasks. The project direction is to investigate how these functions could be connected in one workflow: assessment, identification, practice, reassessment, and progress tracking. Systematic and scoping reviews indicate that many AI-based speech-therapy tools remain fragmented or insufficiently evaluated as complete human-centered systems [14]–[16].

C. Personalized Practice

There is a need to connect identified pronunciation weaknesses to targeted practice activities. The system should be designed to answer three practical questions: which sound or word may need more practice, what activity should be presented, and how the user's later attempt compares with the earlier one. Personalized pronunciation correction, phonemic feedback, computer-assisted pronunciation training, and interactive practice research provide promising building blocks for this direction [4], [10], [17], [18].

D. Longitudinal Progress Monitoring

Many research studies report results from a particular experiment or session. A user-facing practice application may also need to preserve records across days. Potential fields include practice date, target sound or word, number of attempts, correct and incorrect results, score, and previous performance. These records could support a history view, but any claimed improvement would require reliable and comparable measurements. Studies of SayBananas! and Apraxia World illustrate the importance of engagement and repeated home practice in technology-assisted speech intervention [19], [20].

E. Indian-Language and Child-Speech Data

High-quality labeled datasets representing Indian languages, accents, age groups, and pronunciation variations are important for training and evaluation. Child speech can differ from adult speech, so datasets and testing procedures should appropriately represent the intended users. Tripathi et al. addressed Indian-language ASR adaptation, while Kim et al. and Ahn et al. showed the importance of child-specific speech data for pronunciation-related assessment [5], [8], [9]. Resources such as the AI4Bharat-IndicNLP Corpus are valuable for Indic-language text processing, but additional labeled speech resources are needed for pronunciation assessment [12].

F. Reliability and Clinical Boundaries

AI-generated feedback can be incorrect or uncertain. The planned system should avoid presenting an automated score as a definitive diagnosis. Confidence thresholds, validation with suitable data, clear user messaging, and appropriate professional oversight should be considered. Pediatric AI reviews and automatic assessment studies emphasize the need to evaluate reliability, stakeholder involvement, and clinical applicability rather than treating automated outputs as inherently authoritative [2], [9], [14], [15].

VIII. Proposed Research Direction (Planning Only)

Based on the reviewed literature and identified gaps, this paper outlines an initial framework for an AI-assisted multilingual speech-practice web application. At this stage, the team has not built or tested the system. The following workflow describes the intended concept and may change after feasibility analysis.

Proposed workflow: Registration/Login → Language Selection → Initial Pronunciation Assessment → Speech Recording → Speech Preprocessing → AI Speech Analysis → Possible Weak-Sound/Word Identification → Assessment Report → Targeted Practice → Re-analysis → Feedback → Database Storage → Day-wise Progress Dashboard.

A. User Registration and Language Selection

A user would create an account and select a language. The initial scope under consideration is Hindi and English. The architecture may be planned to allow additional Indian languages later, subject to available data and evaluation.

B. Initial Assessment and Recording

The application may present predefined speech sounds, letters, or representative words. The user would record their pronunciation through a device microphone. The assessment content and recording procedure would need to be designed carefully for the intended age group and selected languages.

C. Preprocessing and AI Analysis

Recorded audio may require format handling and preprocessing such as noise reduction, silence removal, normalization, segmentation, or sampling-rate conversion. Possible AI components include ASR, phoneme recognition, acoustic feature extraction, speech-sound classification, pronunciation similarity, and confidence estimation. The final choice of model has not yet been made.

D. Identification and Report

The system is intended to identify sounds or words that may require additional practice and present a simple report. Potential report fields include sounds pronounced as expected, sounds requiring practice, related words, a score, and suggested practice areas. Because automated judgments can be uncertain, the report should avoid diagnostic language.

E. Targeted Practice and Reassessment

The planned practice module could present words, pictures, audio examples, or other activities related to the identified target sound. After practice, the user could record the target again. The system could compare the new attempt with an earlier attempt, provided that the chosen assessment method produces sufficiently reliable and comparable results. Existing work on pronunciation correction, interactive practice, and

digital games supports investigating this kind of repeated practice loop [4], [18]–[20].

F. Progress History

A database could store user details, assessment records, practice attempts, scores, dates, and target sounds or words. A dashboard could display daily scores, completed exercises, practice frequency, and historical results. Any percentage or trend shown should be clearly defined and should not imply clinical improvement unless validated.

G. High-Level Architecture

The proposed architecture can be divided into seven conceptual layers: (1) user interface for registration, language selection, assessment, practice, and dashboard; (2) speech input for microphone access and recording; (3) speech processing for preprocessing and feature extraction; (4) AI assessment for recognition and pronunciation analysis; (5) personalization for selecting practice content; (6) database for user and practice records; and (7) dashboard for displaying current and historical information. These are planned modules, not completed software components.

H. Initial Scope and Feasibility

To keep the project manageable, the team should first determine which speech prompts, languages, audio-processing methods, and evaluation measures can realistically be supported. A limited prototype may be considered after the review and requirements analysis. The project should not claim that it can detect every speech disorder or support every language. The initial objective is to investigate a focused practice workflow and document its limitations.

IX. Challenges and Limitations

Several challenges may affect a future implementation. Speech varies by accent, dialect, speaking speed, voice characteristics, age, and pronunciation style. Home recordings may contain background noise, and microphones differ in quality. Indian languages have distinct phonetic systems, so language-specific evaluation is necessary. Training or adapting models may require labeled datasets that are not readily available. Child speech requires particular attention because adult-trained models may not generalize [5], [8], [9]. AI errors could lead to inappropriate feedback, so the system would need validation and clear communication of uncertainty. The application should be positioned as an assistive practice tool and not as a replacement for qualified speech-language professionals. These limitations should guide development and evaluation rather than be treated as solved problems.

X. Future Research Directions

Possible future directions include adding Indian languages beyond the initial Hindi-English scope; collecting or identifying balanced datasets across ages, accents, and dialects; investigating real-time feedback; and exploring phoneme-level assessment rather than relying only on word recognition. Articulatory analysis may be investigated as an advanced component. Adaptive practice, mobile access, cloud deployment, explainable feedback, and optional professional review are also possible extensions. Each extension would require separate feasibility and evaluation work.

XI. Expected Contribution at the Proposal Stage

At the current stage, the contribution is a proposed research direction and a planned system workflow—not a working application, measured performance, or validated therapy outcome. The intended contribution is to explore an integrated framework connecting multilingual assessment, possible weak-sound or word identification, targeted practice, reassessment, and day-wise history. The central idea is the integration of these stages, while the actual technical novelty and usefulness can only be established after implementation and evaluation.

XII. Conclusion

This review considered AI-assisted speech technologies relevant to multilingual pronunciation practice. The selected literature demonstrates progress in automatic speech analysis, personalized pronunciation correction, multilingual ASR, digital speech practice, and articulatory speech analysis [2], [4]–[6]. It also highlights challenges related to phoneme complexity, language coverage, data availability, evidence quality, and system integration [14]–[16].

The review identifies a potential gap in connecting assessment, identification of practice needs, personalized exercises, reassessment, and longitudinal progress monitoring in one accessible platform. In response, an initial framework for an AI-assisted multilingual speech-practice application has been outlined. The project remains at the planning stage: no implementation, experiment, or result is claimed in this paper. Future work should refine the requirements, verify the selected literature, assess available datasets and models, and then develop and evaluate a limited prototype. Any eventual system should communicate uncertainty and remain an assistive practice tool rather than a substitute for professional clinical assessment.

References

[1] M. S. Remya, R. Raman, R. Sankaran, V. Namboodiri, and P. Nedungadi, "Artificial Intelligence for Speech Classification and Enhancement of Speech and Language Disorders: Techniques, Applications, and Future Directions," *IEEE Access*, vol. 13, pp. 177136–177159, 2025, doi: 10.1109/ACCESS.2025.3620114.

[2] M. Carl, E. Rudyk, Y. Shapira, H. Leavy Rusiewicz, and M. Icht, "Accuracy of Speech Sound Analysis: Comparison of an Automatic Artificial Intelligence Algorithm With Clinician Assessment," *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 9, pp. 3004–3021, 2024, doi: 10.1044/2024_JSLHR-24-00009.

[3] N. R. Benway and J. L. Preston, "Artificial Intelligence-Assisted Speech Therapy for /r/: A Single-Case Experimental Study," *American Journal of Speech-Language Pathology*, vol. 33, no. 5, pp. 2461–2486, 2024, doi: 10.1044/2024_AJSLP-23-00448.

[4] R. Lakshminarayanan, A. S. K. Shaik, and A. Balasundaram, "Automated Speech Therapy Through Personalized Pronunciation Correction Using Reinforcement Learning and Large Language Models," *Results in Engineering*, vol. 25, Art. no. 103943, 2025, doi: 10.1016/j.rineng.2025.103943.

[5] K. Tripathi, R. Gothi, and P. Wasnik, "Enhancing Whisper's Accuracy and Speed for Indian Languages through Prompt-Tuning and Tokeniza-tion," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5, doi: 10.1109/ICASSP49660.2025.10887595.

[6] N. R. Benway, S. Tabatabaee, D. Wang, B. Munson, J. L. Preston, and C. Espy-Wilson, "Perceptual Ratings Predict Speech Inversion Articulatory Kinematics in Childhood Speech Sound Disorders," *Journal of Speech, Language, and Hearing Research*, vol. 69, no. 2, pp. 541–561, 2026, doi: 10.1044/2025
_JSLHR-25-00515.

[7] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, pp. 333–339, 2019, doi: 10.1016/j.jbusres.2019.07.039.

[8] T. Ahn *et al.*, "Automatic Speech Recognition (ASR) for the Diagnosis of Pronunciation of Speech Sound Disorders in Korean Children," *Clinical Linguistics & Phonetics*, vol. 39, no. 10, pp. 913–926, 2025, doi: 10.1080/02699206.2024.2387609.

[9] D. H. Kim *et al.*, "Usefulness of Automatic Speech Recognition Assessment of Children With Speech Sound Disorders: Validation Study," *Journal of Medical Internet Research*, vol. 27, Art. no. e60520, 2025, doi: 10.2196/60520.

[10] A. Rustagi, S. Bajpai, N. Kaur, and S. Siddharth, "Dhvani: A Weakly-Supervised Phonemic Error Detection and Personalized Feedback Sys-tem for Hindi," arXiv preprint arXiv:2506.02166, 2025.

[11] Y. Getman *et al.*, "wav2vec2-based Speech Rating System for Chil-dren with Speech Sound Disorder," in *Proc. Interspeech*, 2022, pp. 3618–3622, doi: 10.21437/Interspeech.2022-10103.

[12] A. Kunchukuttan, D. Kakwani, S. Golla, G. N. C., A. Bhattacharyya, M. M. Khapra, and P. Kumar, "AI4Bharat-IndicNLP Corpus: Monolingual Corpora and Word Embeddings for Indic Languages," arXiv preprint arXiv:2005.00085, 2020.

[13] A. Bhardwaj, M. Sharma, S. Kumar, S. Sharma, and P. C. Sharma, "Transforming Pediatric Speech and Language Disorder Diagnosis and Therapy: The Evolving Role of Artificial Intelligence," *Health Sciences Review*, vol. 12, Art. no. 100188, 2024, doi: 10.1016/j.hsr.2024.100188.

[14] Z. Shi, D. Chung, Y. Du, J. Zhang, S. Raina, and M. Mataric, "Is AI Ready to Support Speech Therapy for Children? A Systematic Review of AI-Enabled Mobile Apps for Pediatric Speech Therapy," in *Proc. 24th Interaction Design and Children (IDC)*, 2025, pp. 479–493, doi: 10.1145/3713043.3728841.

[15] M. Dadgar, C. Ennis, K. Mokgosi, and R. Ross, "Artificial Intelligence (AI)-Driven Technologies for Managing Pediatric Speech and Language Therapy: A Scoping Review," *Digital Health*, vol. 11, 2025, doi: 10.1177/20552076251376533.

[16] C. Deka, A. Shrivastava, A. K. Abraham, S. Nautiyal, and P. Chauhan, "AI-based automated speech therapy tools for persons with speech sound disorder: A systematic literature review," *Speech, Language and Hearing*, vol. 28, no. 1, pp. 1–22, 2025, doi: 10.1080/2050571X.2024.2359274.

[17] D. Korzekwa, J. Lorenzo-Trueba, T. Drugman, and B. Kostek, "Computer-Assisted Pronunciation Training—Speech Synthesis Is Al-most All You Need," *Speech Communication*, vol. 142, pp. 22–33, 2022, doi: 10.1016/j.specom.2022.06.003.

[18] S. Hong, X. Briggs, Q. Zheng, Y. Du, J. Xiong, and T. J.-J. Li, "MORA: AI-Mediated Story-Based Practice for Speech Sound Disorder from Clinic to Home," arXiv preprint arXiv:2510.23887, 2025.

[19] S. McLeod, G. Kelly, B. Ahmed, and K. J. Ballard, "Equitable Access to Speech Practice for Rural Australian Children Using the SayBananas! Mobile Game," *International Journal of Speech-Language Pathology*, vol. 25, no. 3, pp. 388–402, 2023, doi: 10.1080/17549507.2023.2205057.

[20] A. Hair, K. J. Ballard, C. Markoulli, P. Monroe, J. McKechnie, B. Ahmed, and R. Gutierrez-Osuna, "A Longitudinal Evaluation of Tablet-Based Child Speech Therapy with Apraxia World," *ACM Transactions on Accessible Computing*, vol. 14, no. 1, Art. no. 3, pp. 1–26, 2021, doi: 10.1145/3433607.