ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 8, Issue 4, April 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2025 IJMRSET | Volume 8, Issue 4, April 2025 | DOI:10.15680/IJMRSET.2025.0804389

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI-Based Speech Therapy Assistant for Malayalam: A Deep Learning Approach for Language-Specific Speech Disorders

Anas, Fathimathu Suhara, Fathima Hasna, Jasina

Department of Artificial Intelligence and Data Science, Royal College of Engineering and Technology, Kerala, India

Sincy V. Thambi

Assistant Professor, Department of Artificial Intelligence and Data Science, Royal College of Engineering and Technology, Kerala, India

ABSTRACT: Speech therapy is a clinical intervention that addresses communication disorders affecting speech production, language comprehension, fluency, and voice quality. While essential for improving quality of life and social participation, speech therapy faces significant challenges including professional shortages, accessibility barriers, engagement difficulties, subjective assessment practices, and limited personalization. These challenges are magnified for low-resource languages like Malayalam, which lacks standardized assessment tools, therapy materials, and trained professionals familiar with its unique phonological features such as distinctive phonemes and geminate consonants. Conventional speech therapy approaches designed for widely spoken languages often fail to address language-specific articulation patterns, creating a critical need for technology-assisted solutions that can overcome geographical barriers, standardize assessment procedures, and preserve linguistic authenticity while providing accessible intervention for Malayalam speakers with speech disorders. This paper proposes an AI-driven speech therapy assistant tailored for Malayalam speakers with speech disorders. The proposed system leverages deep learning techniques to address common articulation errors, such as mispronunciations and fluency issues like stammering. The framework aims to analyse speech patterns, identify errors, and recommend personalized therapeutic exercises. By focusing on Malayalam-specific phonetics, this research bridges the gap in accessible and culturally relevant speech therapy solutions.

KEYWORDS: Malayalam Speech Therapy, Artificial Intelligence, Speech Disorders, Deep Learning, Language-Specific Solutions

I. INTRODUCTION

Speech therapy is a clinical intervention that addresses communication disorders affecting speech production, language comprehension, fluency, and voice quality. While essential for improving quality of life and social participation, speech therapy faces significant challenges including professional shortages, accessibility barriers, engagement difficulties, subjective assessment practices, and limited personalization. These challenges are magnified for low-resource languages like Malayalam, which lacks standardized assessment tools, therapy materials, and trained professionals familiar with its unique phonological features such as distinctive phonemes and geminate consonants. Conventional speech therapy approaches designed for widely spoken languages often fail to address language-specific articulation patterns, creating a critical need for technology-assisted solutions that can overcome geographical barriers, standardize assessment procedures, and preserve linguistic authenticity while providing accessible intervention for Malayalam speakers with speech disorders. This paper proposes an AI-driven speech therapy assistant tailored for Malayalam speakers with speech disorders. The proposed system leverages deep learning techniques to address common articulation errors, such as mispronunciations of ഷ (sha) as സ (sa) and ഭ (bha) as ഫ (pha), as well as fluency issues like stammering. The framework aims to analyse speech patterns, identify errors, and recommend personalized therapeutic exercises. By focusing on Malayalam-specific phonetics, this research bridges the gap in accessible and culturally relevant speech therapy solutions.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025 | DOI:10.15680/IJMRSET.2025.0804389

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Malayalam, a Dravidian language with 56 unique phonemes, poses distinct phonological challenges that are often overlooked by existing speech therapy technologies. Malayalam's complex phonological system, including phonemes like ഴ (zha) and geminate consonants, presents unique challenges that are not addressed by existing tools. These linguistic features, which are integral to the accurate articulation of Malayalam, require specialized attention in speech therapy applications. Existing speech recognition and therapy tools, predominantly designed for widely spoken languages, often overlook these intricacies, leading to ineffective error detection and misalignment in therapy exercises. For instance, the distinct pronunciation of sounds like the retroflex fricative (ഴ) or the doubling of consonants in geminate pairs (e.g.,) cannot be captured by generalized models trained on languages with different phonological structures. Consequently, a lack of adaptation for these specific phonetic attributes results in suboptimal support for Malayalam speakers, necessitating the development of tailored solutions that can address these unique articulatory challenges. Furthermore, the variation in regional dialects within Kerala compounds the issue, as phoneme realization may differ significantly across different areas, requiring a comprehensive approach that accounts for these diversities.

This paper proposes an AI-based speech therapy assistant designed specifically for Malayalam speakers. The system aims to analyse speech for Malayalam-specific articulation errors, provide personalized therapeutic exercises using everyday vocabulary, and offer real-time feedback and progress tracking. The proposed solution is cost-effective, accessible, and tailored to the linguistic nuances of Malayalam, making it a valuable tool for individuals with speech disorders.

Objectives:

The primary objectives of this research are as follows:

- To develop an AI-driven speech therapy assistant specifically designed for Malayalam speakers with speech disorders.
- To construct a comprehensive Malayalam speech corpus with phoneme-level annotations and articulation error classifications.
- To implement deep learning models for detecting Malayalam-specific articulation errors, such as mispronunciations and fluency issues.
- To provide personalized therapeutic exercises and real time feedback to users based on detected speech errors.
- To enhance accessibility and scalability of speech therapy by leveraging AI-driven solutions for remote users.

In the next section, we review existing research on speech therapy, speech recognition, and AI-driven solutions for speech disorders. While speech therapy applications have made significant progress in widely spoken languages, there remains a gap in addressing language-specific challenges for Malayalam. Existing approaches in speech recognition and articulation therapy lack adaptations for Malayalam's unique phonemes, limiting their effectiveness. This section examines prior studies, identifies key limitations, and highlights the need for a tailored AI-based speech therapy solution.

II. RELATED WORK

Existing speech recognition platforms like Google Speech-to-Text and Microsoft Azure Speech Service excel in major languages but struggle with Malayalam's unique phonemes. Commercial applications like Speech Blubs and Articulation Station lack language-specific adaptations, highlighting the need for targeted solutions. Research on language-specific speech therapy has primarily focused on widely spoken languages. For instance, Shahnawazuddin et al. [1] developed a Hindi pronunciation tool, while Thomas and Muttappallymyalil [2] studied phonological processes in Malayalam speaking children. However, these studies did not translate into practical therapeutic tools. Limited research exists on Malayalam speech therapy. Prema et al. [3] achieved 76 % accuracy in Malayalam phoneme recognition using neural networks, but their work lacked therapeutic applications. Nair and Kumar[4] proposed a Hidden Markov Model based framework but did not address real-world pronunciation challenges. The literature reveals a lack of real-world pronunciation solutions, culturally appropriate therapeutic exercises, and integration of speech analysis with exercise generation

In the next section, we describe the proposed methodology for developing an AI-powered speech therapy assistant for Malayalam. The approach involves data collection, preprocessing, feature extraction, and deep learning-based speech analysis. The system is designed to detect articulation errors, recommend personalized exercises, and provide real-time feedback. This section outlines the framework, model architecture, and implementation strategy.

III. METHODOLOGY

In this section, we are going to discuss the methodology adopted for developing an AI-powered speech therapy assistant for Malayalam which includes data collection, preprocessing, feature extraction, and deep learning-based speech analysis. Since no publicly available dataset exists for Malayalam speech therapy, we manually collect and annotate speech data. The methodology consists of data collection, preprocessing, feature extraction, and classification. A deep learning-based framework is employed to analyse speech patterns, identify errors, and recommend personalized therapeutic exercises.

A. Data Collection and Preprocessing

Since there is no publicly available dataset for Malayalam speech therapy, this study involves manual data collection to create a Malayalam speech corpus. The dataset comprises 200+ hours of speech recordings from 100+ participants, covering an age range of 4–65 years and diverse regional dialects across Kerala. Speech pathologists and linguists annotate the dataset with phoneme-level transcriptions, error classifications (e.g., ഷ → സ), and severity ratings.

As there is no standard Malayalam speech therapy dataset, we are manually collecting voice samples from native speakers. The dataset consists of Malayalam alphabets, numbers, sentences, phrases, and words, ensuring comprehensive phoneme representation for speech analysis. It includes recordings from individuals with and without speech disorders, covering a range of pronunciation challenges to ensure diverse and representative data for speech therapy applications. To capture phonetic diversity, the dataset includes Malayalam vowels (അ, ആ, ഇ...ഔ, അം , അഃ) and consonants (ക, ഖ, ഗ...റ), numbers from 1 to 10 (ഒന്ന്, രണ്ട്, മൂന്ന്...), common sentences (e.g., "എനിക്ക് ഒരു കാപ്പി േവണം" – I want a coffee), phrases (e.g., "ക്ഷമിക്കണം" – Sorry), and words (e.g., ,"കുടം" – Pot, "ആകാശം" – Sky). To enhance pronunciation training, the dataset incorporates phonetically similar but distinct sounds, such as Sha-Sa (ശ-സ, ഷസ) – (വിശ്വാസം - വിസ്വാസം Vishwasam - Viswasam, belief), (ശാസനം - സാസനം Shaasanam - Saasanam, decree); Ba-

Pha (ബ-ഫ) – (ബലം - ഫലം Balam - Phalam, strength); Zha-Ya (ഴ-യ) – (മഴ - മയ Mazha - Maya, rain), (വഴി -വയി Vazhi - Vayi, way). Additionally, challenging words with complex phonetic structures, such as മ (forgiveness), സ്തുതി

(praise),ശ്രവണം(listening), and ഗർജ്ജനം (roaring), are included. The dataset also features tongue twisters to assess articulation accuracy. These elements aid in evaluating phoneme transitions and articulation accuracy, enabling an AI driven speech therapy system to analyse speech patterns, identify pronunciation errors, and provide targeted therapeutic recommendations.

The preprocessing step ensures speech clarity and consistency by applying noise reduction, segmentation, normalization, and filtering. Noise reduction is achieved using spectral subtraction techniques, while segmentation isolates phoneme-level speech samples. Normalization maintains uniform amplitude levels, and filtering removes background noise to enhance speech quality. The preprocessing pipeline is as shown in Figure 1

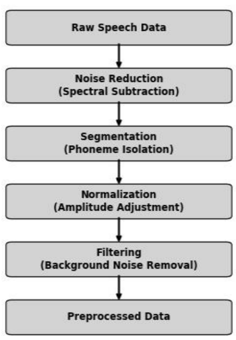

Figure 1: Preprocessing pipeline

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The speech preprocessing pipeline consists of multiple stages to refine raw speech data for accurate analysis and classification, as shown in Figure 1. The process begins with raw speech data, which are recorded from users. The first step in preprocessing is Noise Reduction, where spectral subtraction techniques are applied to minimize background noise and enhance the clarity of speech signals. Next, the segmentation step isolates phonemes, allowing precise analysis of individual speech components. This is followed by Normalization, where amplitude adjustments are made to ensure uniform volume levels across all recordings. The final step is Filtering, which further eliminates any residual background noise, ensuring that only relevant speech signals are retained. After completing these steps, the processed speech data is ready for further feature extraction and classification in the speech therapy system. This structured approach enhances the system's ability to accurately detect pronunciation errors and provide effective feedback for speech therapy applications.

B. Feature Extraction and Classification

To detect articulation errors, the system extracts acoustic features critical to speech analysis. These include Mel Frequency Cepstral Coefficients (MFCCs) for spectral representation [5], prosodic features such as stress and, and articulatory features like formant transitions for phoneme analysis

The deep learning architecture used for classification and error detection consists of a Convolutional Neural Network (CNN) for feature extraction, capturing phoneme structures [6]. A Bidirectional Long Short-Term Memory (BiLSTM) network is employed for temporal modelling, improving context-aware phoneme recognition [7]. Additionally, an attention mechanism is incorporated to focus on critical phonetic transitions, enhancing accuracy in mispronunciation detection.

The model is trained and optimized using data augmentation techniques, including environmental noise addition and pitch variations to improve robustness. Speaker variability training ensures adaptability across different voice characteristics, while context-sensitive training helps refine phoneme recognition in real-world speech scenarios.

In the next section, we present the Proposed System Architecture, which defines the structural framework of our AI-driven speech therapy system. This includes key modules such as data acquisition, preprocessing, feature extraction, model training, and real-time feedback generation. The architecture is designed to efficiently process Malayalam speech data, ensuring precise phoneme recognition and pronunciation analysis.

IV. PROPOSED SYSTEM ARCHITECTURE

In this section, we describe the Proposed System Architecture, which outlines the core components and workflow of the AI-driven speech therapy system. It covers data collection, preprocessing, feature extraction, model training, and real-time speech analysis. The architecture is designed to efficiently process Malayalam speech, ensuring accurate phoneme recognition and pronunciation assessment.The proposed system will consist of five modules: a Speech Input Module to capture user speech through a mobile interface, an Acoustic Analysis Module to analyse speech for errors, an Error Detection Module to identify specific articulation errors, an Exercise Recommendation Module to generate personalized exercises, and a Feedback Module to provide real-time feedback and progress tracking. The proposed system architecture is as shown in Figure 2

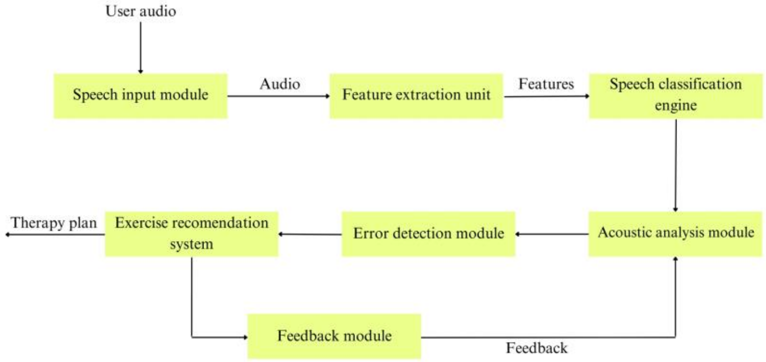

Figure 2: Proposed System Architecture

© 2025 IJMRSET | Volume 8, Issue 4, April 2025 | DOI:10.15680/IJMRSET.2025.0804389

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The proposed system architecture for the AI-driven speech therapy application consists of multiple interconnected modules, as shown in Figure 2. The system begins with the Speech Input Module, which captures user audio for processing. This raw speech data is then passed to the Feature Extraction Unit, where essential features such as pitch, frequency, and phoneme structures are extracted for further analysis. The extracted features are processed by the Speech Classification Engine, which categorizes the speech patterns based on predefined models trained on the Malayalam speech dataset. The classified speech data is then analysed in the Acoustic Analysis Module, where deviations in pronunciation, fluency, and articulation are detected. These deviations are forwarded to the Error Detection Module, which identifies specific phonetic, or articulation errors based on linguistic rules and machine learning algorithms. Once errors are identified, the Exercise Recommendation System suggests personalized speech exercises tailored to the user's needs. These exercises are designed to improve phoneme articulation, pronunciation accuracy, and overall speech fluency. The system then delivers corrective feedback through the Feedback Module, which provides real-time insights, progress tracking, and recommendations for continued practice. This structured approach ensures accurate speech assessment and targeted therapy recommendations, making the system a valuable tool for individuals with speech disorders.

In the next section, we outline the implementation plan, detailing the systematic approach for developing and deploying the proposed system. This includes dataset collection, preprocessing techniques, model selection, training methodologies, and evaluation metrics. The implementation strategy is designed to ensure scalability, real-time performance, and adaptability to various speech disorders, making the system effective for personalized speech therapy.

V. IMPLEMENTATION PLAN

In this section, we present the implementation plan, which details the step-by-step development and deployment of the proposed system. This includes dataset collection, preprocessing techniques, model training strategies, and evaluation methods. The implementation focuses on optimizing system performance, scalability, and adaptability for effective speech therapy applications.

The development of the proposed system will be carried out in four sequential phases, focusing on critical aspects of the project. Each phase will address specific challenges and deliverables, ensuring a structured and systematic approach to system development. The first phase involves data collection and corpus development, including the collection of speech recordings from 100+ participants across Kerala, covering diverse age groups and regional dialects. The dataset will be annotated with phoneme-level transcriptions and error classifications, followed by preprocessing to ensure quality and consistency. The second phase focuses on model development, where deep learning models will be designed and trained for feature extraction and error detection. The model will be optimized for Malayalam specific phonemes and articulation patterns and validated using a subset of the collected data. The third phase involves exercise development and system integration. Therapeutic exercises tailored to common Malayalam speech errors will be developed and integrated with the feedback and progress tracking system. These exercises will include auditory discrimination tasks, articulatory guidance, and fluency enhancement techniques. Preliminary testing will be conducted with a small group of users. The fourth phase focuses on testing and refinement. Extensive user testing will be performed with individuals with speech disorders and speech therapists. The system will be refined based on feedback and performance metrics and prepared for deployment in real-world settings.

VI. CONCLUSION

This paper proposes an AI-based speech therapy assistant for Malayalam speakers, addressing a critical gap in accessible and culturally relevant solutions. By leveraging deep learning and focusing on real-world pronunciation challenges, the system aims to improve communication abilities and quality of life for individuals with speech disorders. The proposed phased implementation ensures a structured and systematic development process, with future work focusing on dialect adaptation and broader language support.

The novelty of this paper lies in the development of the first manually curated Malayalam speech therapy dataset, addressing the lack of standardized resources for speech disorder assessment in the language. Unlike existing datasets focused on general speech recognition, our dataset is specifically designed for phoneme-level analysis, incorporating annotated pronunciation variations, error classifications, and severity ratings provided by speech pathologists. Additionally, the dataset includes a diverse set of phonetic challenges, such as confusable phoneme pairs (Sha-Sa, Ba-Pha, Zha-Ya, etc.), tongue twisters, and complex words, which are critical for evaluating articulation accuracy. The

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

proposed deep learning model integrates CNNs, Bi-LSTMs, and attention mechanisms to improve phoneme recognition and speech error detection, making this system the first AI-driven Malayalam speech therapy tool with context-sensitive feedback and personalized recommendations.

REFERENCES

[1] S. Shahnawazuddin et al.," Pronunciation learning in children: A developmental approach using computer games," Speech Communication, 2018.
[2] S. Thomas and M. Muttappallymyalil," Phonological processes in typically developing Malayalam-speaking children," Language in India, 2019.
[3] N. Prema et al.," Malayalam speech recognition using wavelet features and neural network," International Journal of Speech Technology, 2017.
[4] S. Nair and P. Kumar," Malayalam speech analysis for articulation disorders using hidden Markov models," International Journal of Advanced Research in Computer Science, 2017.
[5] Thulasiram Prasad, P. (2024). A Study on how AI-Driven Chatbots Influence Customer Loyalty and Satisfaction in Service Industries. International Journal of Innovative Research in Computer and Communication Engineering, 12(9), 11281-11288.
[6] S. Davis and P. Mermelstein," Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357–366, 1980.
[7] Y. LeCun, Y. Bengio, and G. Hinton," Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[8] A. Graves and N. Jaitly," Towards end-to-end speech recognition with recurrent neural networks," Proceedings of the 31st International Conference on Machine Learning (ICML), pp. 1764–1772, 2014.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com