Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Echo Learn: An AAC-Based Assistive App for Autism and Down Syndrome

1st Rakshith P
Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru,India
Email: rakshith@jssstuniv.in

2nd Deepika M
Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India Email: murthydeepika611@gmail.com

3rd Suhas M
Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India
Email:suhasmanishshetty@gmail.com

4th Nagaveni L S
Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru,India
Email: nagavenils59@gmail.com

5th Sanjana M C
Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru,India
Email:sanjanamc01@gmail.com

***Abstract*— Some children find it difficult to speak clearly and communicate their needs. They may use AAC tools to form sentences using symbols, but these tools do not help them improve how they pronounce words. Because of this, children may continue making the same mistakes in speech without understanding how to correct them. This project presents Echo Learn, a web-based system designed to support both communication and basic speech practice. The application allows children to select symbols to form sentences and also try speaking simple words or sentences. The spoken input is checked using a speech analysis service, which identifies where mistakes occur in pronunciation. To help the child improve, the system provides simple visual hints that show how to move the lips and tongue while speaking a particular sound.. The system also saves results from each attempt and shows progress in an easy format. Overall, the proposed system provides a simple and practical way to support speech improvement along with communication in a single platform.**

Keywords— AAC, Speech Practice, Pronunciation Error, Assistive System, Speech Support, Web-Based Tool, Learning Aid.

I. INTRODUCTION

Clear speech communication plays a very important role in everyday human interaction and social development. For children with neurodevelopmental conditions, articulation disorders, or other speech impairments, developing functional communication skills is often a long-term and ongoing challenge. These children frequently depend on Augmentative and Alternative Communication (AAC) systems to express basic needs and to interact with their environment.

Although AAC platforms have been helpful in supporting expressive communication, they mainly focus on enabling output through symbol selection and text-to-speech mechanisms and do not fully address the underlying pronunciation difficulties that many users experience. Traditional speech therapy usually requires regular sessions with a therapist, fixed schedules, and subjective evaluation of progress. These methods are time-consuming, difficult to extend into home settings, and give parents limited objective data about how their child is actually improving.

A major gap in current assistive technology is the absence of visual articulation guidance. Children with speech difficulties cannot easily correct sounds they cannot see or understand physically. Simply knowing that a sound is wrong is not enough; they also need to know where the tongue should go, how the lips should be shaped, and how the breath should flow. This visual aspect of articulation learning is almost entirely missing from existing AAC and pronunciation-training platforms.

Recent advances in cloud-based speech recognition have made it possible to perform phoneme-level pronunciation assessment within standard web architectures, allowing real-time detection of specific articulation errors without the need for specialized hardware. This paper presents Echo Learn, a web-based assistive application that combines AAC communication support with an integrated pronunciation learning module. The system offers symbol-based sentence construction, phoneme-level error detection using the Microsoft Azure Pronunciation Assessment API, and a dedicated articulation visualization system that shows lip configuration, tongue position, and airflow direction for each target phoneme. Session-based progress tracking and graphical analytics provide clear, objective visibility into pronunciation development over time.

The main contributions of this work are:

1. a unified platform that integrates AAC communication and structured pronunciation practice;
2. real-time phoneme-level error detection and single-error-first corrective feedback;
3. an articulation visualization system that demonstrates lip shape, tongue position, and airflow guidance for each phoneme; and
4. session-based analytics that enable objective monitoring of pronunciation improvement trends.

II. MOTIVATION AND PROBLEM DEFINITION

Children with speech and communication difficulties form a diverse group with complex and varied needs. While AAC platforms successfully meet immediate communication demands,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

they do not actively support the deeper goal of improving a child's ability to produce sounds accurately. This gap between communication assistance and pronunciation development represents a significant unmet need in current assistive technology.

Between therapy sessions, children and caregivers often lack practical tools that can provide clear, real-time feedback on pronunciation attempts. Without this kind of ongoing support, incorrect articulation patterns can persist for long periods without correction. A particularly important limitation is the absence of visual articulation guidance in existing systems. Pronunciation correction is not just about noticing that a sound is wrong; it is about understanding the exact physical mechanism of correct sound production. Lip position, tongue placement, and airflow direction form the physical basis of articulation, yet no current AAC platform offers this kind of visual feedback.

The key problems addressed in this work are:

- existing AAC systems provide no structured pronunciation practice or phoneme-level correction;
- there is no visual articulation guidance in current AAC platforms that shows lip shape, tongue position, and airflow for specific sounds;
- children do not have access to objective, real-time feedback on specific articulation errors between therapy sessions;
- caregivers lack data-driven insight into pronunciation progress or recurring weak sound patterns over time; and
- communication support and pronunciation development remain separate functions with no integrated platform that addresses both together.

Echo Learn addresses these limitations by providing a single, accessible platform where a child can communicate using AAC symbols and at the same time receive structured, phoneme-specific pronunciation guidance with visual articulation support.

III. LITERATURE REVIEW

Augmentative and Alternative Communication systems have been extensively studied and deployed to support individuals with speech and communication impairments [8]. Established platforms such as Tobii Dynavax and Proloquo2Go enable expressive communication through image-based symbol selection and text-to-speech output [9]. These systems significantly improve communication ability but do not incorporate structured pronunciation assessment or articulation improvement mechanisms [10].

Research in automatic pronunciation evaluation has progressed considerably [3]. Early approaches relied on rule-based phoneme matching and word-level scoring, lacking the granularity required for precise identification of individual articulation errors [3]. Subsequent work introduced phoneme-level evaluation using hidden Markov models and acoustic feature analysis, enabling more detailed mispronunciation detection [5]. These advances improved feedback specificity but remained largely confined to research environments [5].

The emergence of deep learning-based speech recognition has significantly enhanced pronunciation assessment capabilities [6]. Cloud-based services now expose phoneme-level accuracy scoring, error classification, and prosody evaluation through accessible APIs [1]. Microsoft Azure Pronunciation Assessment provides per-phoneme accuracy scores, error type classifications including omission, insertion and mispronunciation, and word-level metadata through a standard API interface [1].

Children's speech presents additional complexity due to higher fundamental frequency, variable articulation patterns, and developing phonological systems that differ substantially from adult speech [2]. Pronunciation assessment accuracy for children remains an active area of development with limited practical deployment in assistive technology contexts [11].

Visual articulation guidance has been demonstrated as an effective complement to auditory feedback in pronunciation learning [10]. Research indicates that combining visual mouth position demonstrations with auditory correction improves articulation outcomes, particularly for learners who struggle to internalize sound production from audio alone [10]. Despite this evidence, integration of lip shape, tongue position, and airflow visualization within AAC platforms remains almost entirely absent from existing implementations [7]

A clear research gap exists in combining AAC communication support, phoneme-level pronunciation assessment, visual articulation guidance showing lip configuration, tongue position and airflow, and session-based analytics within a unified web-based platform. Echo Learn addresses this gap directly.

A comparative analysis of existing AAC systems and Echo Learn is presented in TABLE I

TABLE I
COMPARISON OF ECHOLEARN WITH EXISTING AAC SYSTEMS

Feature	Tobii Dynavox	Prolo-quo2Go	Traditional AAC	Echo Learn
Communication Support	Yes	Yes	Yes	Yes
Phoneme-Level Feedback	No	No	No	Yes
Articulation Visualization	No	No	No	Yes
Tongue and Airflow Guidance	No	No	No	Yes
Progress Analytics	No	No	No	Yes
Web-Based Architecture	No	No	No	Yes

IV. PROPOSED SYSTEM OVERVIEW

Echo Learn is designed as a unified web-based assistive application that combines two core functional modules—an AAC communication module and a pronunciation practice module—within a single child-facing interface accessed through a dedicated child login.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

A. AAC Communication Module

The AAC module provides a symbol-based communication board organized into categories such as needs, emotions, actions, objects, and social expressions. Each tile, when selected, speaks the corresponding word aloud using the Web Speech API and adds it to a sentence construction bar The child creates sentences by adding words one at a time and then triggers the full sentence to be spoken using a dedicated speak button. Some commonly used response tiles work as direct-speak buttons that skip the category navigation completely. The board also allows tiles to be added or removed, so that the layout can be adjusted to match each child's specific communication needs.

B. Pronunciation Practice Module

The pronunciation practice module provides structured speech practice through word- and sentence-based exercises that are grouped by difficulty level. The child attempts to reproduce a target word or sentence by speaking into the microphone. The Microsoft Azure Pronunciation Assessment performs phoneme-level analysis, and the system identifies the most problematic phoneme in that attempt.

A central contribution of Echo Learn is its articulation visualization system. When a phoneme error is detected, the system displays a focused visual showing the correct lip configuration, tongue position, and airflow direction for that specific sound. This helps children understand not only that a sound is incorrect but also exactly how to produce it physically—where to place the tongue, how to shape the lips, and how to direct the breath. The visual is shown alongside a simple, child-friendly correction hint in encouraging language. The child retries the full word or sentence, with up to five attempts allowed before the system advances with a positive message and no negative reinforcement.

C. Design Principles

Echo Learn is built around three guiding principles: integration (AAC and pronunciation practice in one platform), specificity (single-phoneme correction per attempt with precise visual articulation guidance), and accessibility (child-appropriate language, visual feedback, and simple interaction throughout).

V. SYSTEM ARCHITECTURE

Echo Learn follows a modular client-server architecture built on the MERN stack, comprising React.js on the frontend, Node.js and Express.js on the backend, and MongoDB as the database, integrated with the Microsoft Azure Pronunciation Assessment API. The overall system architecture of Echo Learn is illustrated in Fig. 1.

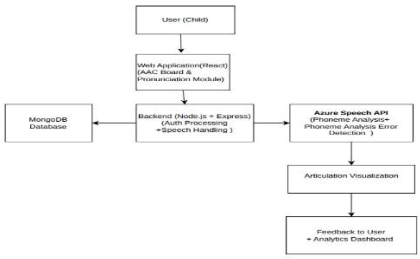

Fig. 1: System Architecture Workflow

A. Authentication Module

The authentication module manages child login verification. Credentials are checked on the Express.js backend by comparing them with the data stored in MongoDB, which connects each child's practice history and progress records to their individual profile across different sessions.

B. AAC Interaction Module

The AAC module handles the symbol-based communication board, tile selection, sentence buffer construction, and speech synthesis via the Web Speech API using a locally defined pronunciation mapping table that corrects Indian-English TTS inconsistencies

C. Pronunciation Assessment Module

The pronunciation assessment module captures microphone input via the Media Recorder API and transmits the audio to the Node.js backend as a binary blob along with the reference text. The backend forwards both to the Microsoft Azure Pronunciation Assessment API, configured at phoneme-level granularity. The returned JSON response is parsed to identify the phoneme with the lowest accuracy score, which is mapped to the articulation visualization system to return lip configuration, tongue position, airflow guidance, and a correction hint for display. The detailed pronunciation assessment workflow is shown in Fig. 2.

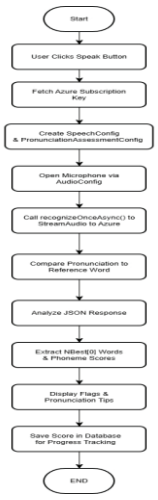

Fig. 2: Pronunciation Assessment Workflow

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

D. Articulation Visualization System

A structured phoneme-to-articulation mapping system associates each Azure-returned phoneme identifier with a dedicated visual that demonstrates the correct lip shape, tongue position, and airflow direction for that sound. This system constitutes a core technical contribution of Echo Learn, providing children with precise physical guidance for sound correction that text hints alone cannot convey.

E. Analytics Module

Processes stored attempt history from MongoDB to compute average phoneme accuracy across sessions and per-item score progression over sequential attempts. Results are rendered as a phoneme accuracy bar chart and a progress line chart using the Recharts library. The core system components and technologies used in Echo Learn are summarized in Table II.

TABLE II
ECHOLEARN SYSTEM COMPONENTS

Component	Technology	Description
Frontend	React.js	AAC board, practice UI, analytics
Backend	Node.js, Express.js	REST API, speech processing
Database	MongoDB	Profiles, attempt history
Speech Assessment	Azure Pronunciation Assessment	Phoneme scoring and error detection
Articulation Visualization	Local mapping system	Lip, tongue, airflow guidance per phoneme
Speech Synthesis	Web Speech API	Reference pronunciation playback
Analytics	Recharts	Bar chart and line chart rendering

VI. METHODOLOGY

The EchoLearn system operates through two integrated workflows — AAC communication and pronunciation practice — supported by a session-based analytics layer.

A. AAC Communication Workflow

Communication tiles are organized into predefined categories and rendered as a symbol-based grid. Each tile selection triggers speech synthesis via the Web Speech API using a locally defined pronunciation mapping table constructed to produce accurate phonetic output under Indian-English TTS conditions. Selected words are appended sequentially to a sentence buffer. Full-sentence playback is initiated through a dedicated speak control. High-frequency response tiles are configured as direct-speak elements that bypass category navigation to minimize interaction steps. The AAC communication process flow is illustrated in Fig. 3.

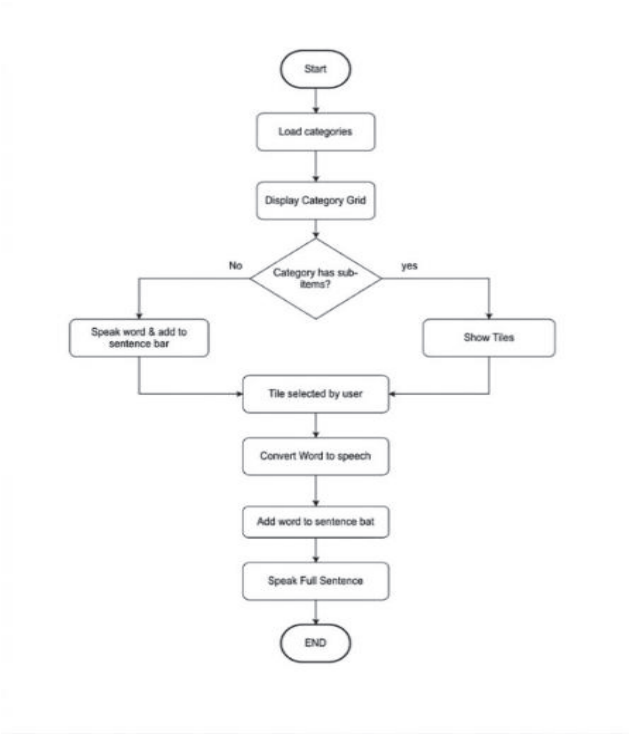

Fig. 3: AAC Communication Workflow

B. Dataset Structure

The word dataset is categorized by communication domain and difficulty level, determined by syllable count and phoneme complexity. The sentence dataset consists of a fixed set of essential daily-life communication phrases selected for practical relevance to the target user group, organized by category into needs, health, feelings, and social expressions.

C. Pronunciation Assessment

Audio is captured via the MediaRecorder API and transmitted to the Node.js backend as a binary blob along with the reference text. The backend forwards both to the Microsoft Azure Pronunciation Assessment API, configured at phoneme-level granularity with miscue detection enabled. The API returns per-phoneme accuracy scores in the range, along with error-type classifications (mispronunciation, omission, insertion) and word-level metadata.

D. Phoneme Error Identification and Articulation Guidance

The system reads the returned response and picks out the phoneme that has the lowest accuracy score in that attempt. This phoneme is treated as the sole correction target for the current feedback cycle. The identified phoneme is matched with the articulation visualization system, which then returns a visual that shows the correct lip shape, tongue position, and airflow direction for that particular sound. This visual is shown together with a simple, child-friendly correction hint. By giving only one correction per attempt, along with clear physical articulation guidance, the system keeps the child's mental load low and focuses directly on the specific way the sound is being produced, instead of offering general or vague feedback.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

E. Retry Evaluation

Each attempt is evaluated with a new call to the Azure API, independent of earlier attempts. This way, the feedback is based only on the child's current speech and does not carry forward any earlier errors or history. A maximum of five attempts is permitted per practice item. If all attempts are exhausted, the system advances automatically with an encouraging message and no negative reinforcement.

F. Progress Tracking and Analytics

Each attempt is persisted in MongoDB, including the child identifier, target item, attempt number, overall accuracy score, per-phoneme scores, error classifications, articulation guidance delivered, and a timestamp. The average phoneme accuracy across sessions is computed as:

$$\text{AvgPhonemeScore}_p = \left(\sum_{i=1}^{N} \text{Score}_{p,i}\right) / N$$

where $\text{Score}_{p,i}$ is the accuracy score of phoneme p in attempt i, and N is the total evaluated attempts for that phoneme. The system generates two visualizations—a phoneme-accuracy bar chart and a progress line chart—rendered using the Recharts library within the React frontend.

VII. IMPLEMENTATION DETAILS

The system is built as a web-based client–server application using the MERN stack. The overall design is kept modular so that the code stays organized, easy to understand, and simple to extend when new features are added. The frontend is developed with React.js, the backend runs on Node.js with Express.js, and all data is stored in MongoDB using Mongoose as the object-document mapper. The system also connects to the Microsoft Azure Pronunciation Assessment API to evaluate speech and give feedback to the user.

The frontend handles the user interface, including the AAC communication board, the pronunciation practice controls, the articulation visualization, and the analytics dashboards. React components are grouped around main features, so there is one section for the symbol board, another for building sentences, one for audio recording, one for showing feedback, and one for displaying progress charts. Basic React hooks such as useState, useEffect, and useContext are used to manage the application state across different parts of the interface, which helps keep the screen updated as the child selects tiles, records speech, or checks progress. This makes it possible to match the UI with the child's current activity—for example, updating the sentence bar when tiles are picked, showing feedback right after each pronunciation attempt, and refreshing the analytics charts when the progress page is opened. The interface is kept simple and child-friendly, with large tiles, clear icons, and very little text, so that children can use the system with little or no adult help.

The backend runs on Node.js with the Express.js framework and provides API endpoints for different tasks. When a child logs in, the backend checks the entered credentials against the records stored in MongoDB and links the session to that child's profile. Separate routes handle AAC-related actions such as loading tiles and saving custom layouts, pronunciation-practice operations such as sending audio for assessment and storing attempt results, and analytics requests such as retrieving score history for a particular phoneme or word. The modular design of these routes helps separate one part of the system from another, so that changes in one module—for example, adding a new visualization or changing how scores are computed—do not affect the rest of the system. The backend also includes basic error handling and simple logging to help spot and fix problems during testing and while preparing the system for deployment.

Speech pronunciation evaluation is handled using the Microsoft Azure Cognitive Services Pronunciation Assessment API. When the child speaks into the microphone, the browser records the audio using the MediaRecorder API and sends the captured file as a binary blob to the Node.js backend. Along with the audio, it also transmits the matching reference text for that word or sentence so the backend can pass both to the pronunciation-assessment service for analysis. The backend forwards both the audio and the reference text to Azure, which returns a JSON response with overall accuracy, per-phoneme scores, error types (such as mispronunciation, omission, or insertion), and word-level information. The backend processes this response to find the phoneme with the lowest accuracy score, which is treated as the main sound that needs correction.When the child finishes speaking, the system prepares a clear and simple response that includes the overall pronunciation score, highlights the problematic phoneme, and adds any extra details the frontend needs to display feedback and update the analytics. In the React frontend, the articulation visualization system is built as a basic JavaScript mapping object. Each phoneme identifier received from Azure is linked to a specific visual element that shows how the lips and tongue should move and how the airflow should flow to produce that sound correctly. For example, the "th" sound is linked to a picture where the tongue is placed between the teeth, the lips are slightly open, and the airflow goes through the small gap. When the backend returns the lowest-scoring phoneme, the frontend looks it up in this mapping and retrieves the matching visual and a short, child-friendly hint (such as "bite your tongue" or "lips together, then release"). This visual-feedback pair is shown in the feedback area, giving the child a clear physical idea of how to correct the sound instead of just hearing the right audio.

The database layer uses MongoDB, with Mongoose to define schemas and manage data operations. Child profiles are stored as individual documents that include basic details such as the child's name, age, and login information. Practice sessions and pronunciation attempts are stored as nested documents under each child's profile, so that related data stays grouped together. Each attempt record includes the child identifier, the target item (word or sentence), the attempt number, the overall score, an array of per-phoneme scores, error classifications, the articulation guidance shown, and a timestamp.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

This structure keeps the data rich and detailed, while still simple to query and work with. For instance, the analytics engine can quickly retrieve all pronunciation attempts made by a specific child or for a particular word, then calculate averages, track progress over time, or extract other useful metrics from that dataset.

The analytics engine processes the stored attempt history and turns it into meaningful summaries. On the backend, the system reads the attempt records from MongoDB, groups the scores by phoneme and by item, and calculates averages across sessions. For each phoneme, it finds the average accuracy across all attempts and marks any phonemes that fall below a set threshold as weak sounds that need extra practice. The engine also watches how the scores for individual words or sentences change over time, which helps identify gradual improvement or ongoing difficulties. The computed results are sent to the frontend as a simple JSON object that can be easily displayed using the Recharts library. The frontend shows the results either as a bar chart of phoneme-level accuracy or as a line chart of progress across attempts, depending on which view the user picks.

Through this mix of interactive frontend design, backend processing, cloud-based speech assessment, and a flexible database structure, Echo Learn delivers a practical and usable platform for pronunciation practice and communication support. The implementation is modular enough to allow future additions, such as new visualizations, offline models, or support for different languages. At the same time, the current design keeps the interface simple and focused so that children can follow the guidance without getting confused by technical details.

VIII. RESULTS AND PERFORMANCE EVALUATION

The system was tested through multiple practice sessions under conditions that loosely mimic real user interaction. The focus was on checking whether the platform could correctly record pronunciation scores, identify weak words, compute analytics efficiently, and update the visualizations in a timely way. The user interface and system outputs are shown in Figs. 4–8.

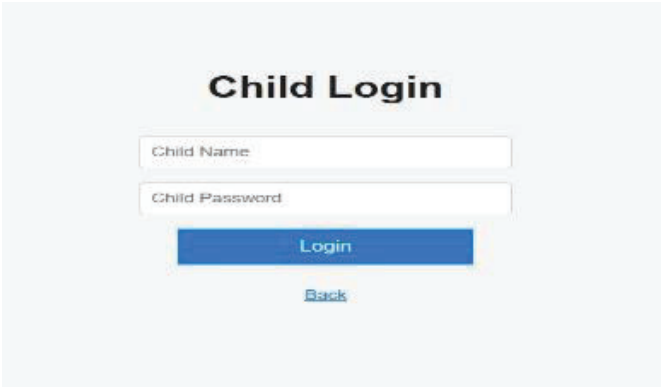

Fig. 4: Child Login Interface

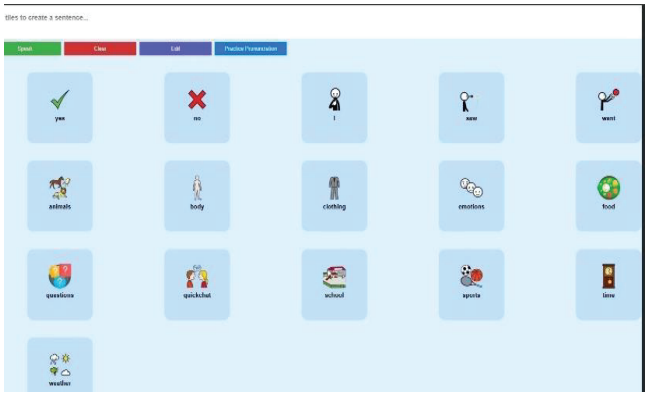

Fig. 5: AAC Communication Board

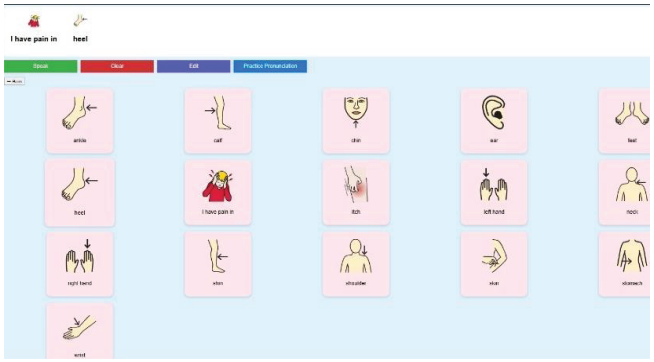

Fig. 6: Category Tile Selection and Sentence Construction showing "I have pain in the heel"

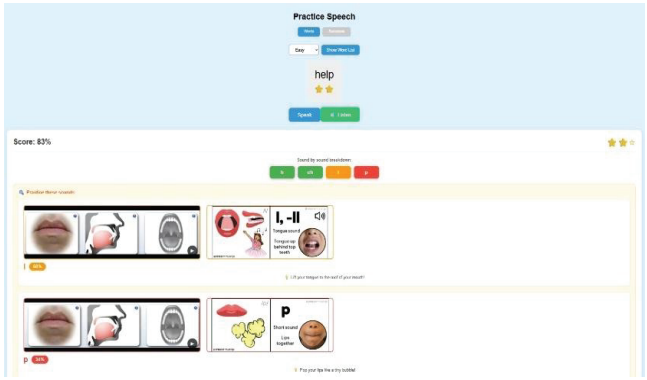

Fig. 7: Pronunciation Practice with Articulation Visualization

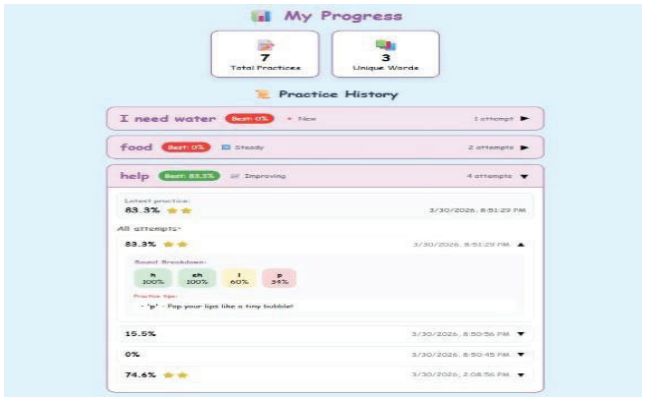

Fig. 8: Progress Tracking Dashboard

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

A. AAC Module Evaluation

The AAC communication board worked correctly in all predefined categories. Tile selection, sentence building, and playback operated without noticeable delay. The locally defined pronunciation mapping table successfully corrected several Indian-English TTS quirks, so that even commonly mispronounced words sounded closer to standard speech. Tiles that were set as high-frequency response elements functioned as direct-speak buttons and did not require navigating through category menus, which sped up the communication process for common phrases. Adding and removing tiles also worked reliably, allowing basic customization of the board for different children.

B. Pronunciation Assessment Results

Table III summarizes pronunciation assessment results obtained from practice sessions in both word and sentence modes.

TABLE III
PRONUNCIATION ASSESSMENT RESULTS

Item	Type	Difficulty	Score (%)	Weak Phonemes
food	Word	Easy	82.8	—
help	Word	Easy	82.7	—
bathroom	Word	Hard	70.6	b, ae, th
breakfast	Word	Hard	63.4	b, r,eh, k, f
I need help	Sentence	Medium	52.5	d, h, eh, p
Please call for help	Sentence	Hard	58.1	p, l, iy, k, aa

Easy-level words generally scored above 80%, which suggests that simpler vocabulary was often produced with relatively clear articulation, especially when visual guidance was given. Hard-level words and multi-word sentences scored lower, reflecting the increased difficulty of longer or more complex utterances. The articulation visualization system delivered the correct lip configuration, tongue position, and airflow direction for each identified weak phoneme, giving children specific physical cues instead of only general feedback

C. Per-Phoneme Accuracy Analysis

The analytics module aggregated phoneme scores across all sessions and produced a phoneme-accuracy bar chart. The chart showed that certain phonemes such as uw, ao, r, and f tended to stay above 75% accuracy, indicating that children were usually able to produce them correctly. Other phonemes such as b, ae, l, aa, and k often scored below 40%, marking them as clear targets for focused articulation-guided practice. This breakdown at the phoneme level helps pinpoint exactly which sounds are causing the most trouble, which a simple word-level score would not reveal.

D. Progress Tracking Evaluation

Table IV shows how the score changed for the word bathroom over five successive attempts.

TABLE IV
PROGRESSIVE ATTEMPT SCORES — BATHROOM

Attempt	Score (%)	Weak Phoneme	Articulation Guidance Delivered
1	70.6	b	Lip closure and release
2	73.2	ae	Wide mouth open position
3	76.8	th	Tongue between teeth airflow
4	79.1	th	Tongue between teeth airflow
5	82.4	r	Tongue curl back position

In each attempt, the system identified the weakest phoneme and showed the corresponding visual articulation guidance. This allowed the child to gradually improve the production of the word as the focus shifted from one sound to another. By the fifth attempt, the overall score had increased from 70.6% to 82.4%, which suggests that repeated, guided practice can lead to measurable improvement.

E. System Performance

The system remained stable during testing. Azure API calls returned responses quickly, so feedback appeared almost in real time without obvious lag. AAC interactions, sentence construction, and speech playback were smooth, with no noticeable freezing or delay. The analytics dashboard loaded correctly and updated charts as soon as the user switched to the progress view. Overall, the platform demonstrated that it can reliably record and process pronunciation data, deliver targeted articulation feedback, and show progress in a clear, visual way.

IX. ADVANTAGES AND LIMITATIONS

A. Advantages

Echo Learn offers several important improvements over existing AAC platforms. By combining AAC communication and structured pronunciation practice in a single web-based application, it reduces the need for separate tools and keeps everything in one place for the child, parents, and therapists. The use of Microsoft Azure Pronunciation Assessment at the phoneme level allows very specific identification of which sounds are being misarticulated, something that word-level scoring cannot do.

The articulation visualization system is one of the main strengths of the platform. By showing how the lips move, where the tongue should be placed, and how the airflow should flow for each phoneme, it helps children understand the physical process of sound production, not just the sound itself. The single-error-first feedback strategy keeps instructions simple and avoids overwhelming the child with too many corrections at once. Each attempt is checked independently, so the feed-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

back always reflects the child's current articulation. The session-based analytics also give caregivers a clear, visual record of how pronunciation is changing over time, which can be useful for planning further practice. Finally, the web-based MERN stack allows the system to run on multiple devices without installing special software, making it easier to use at home or in a classroom.

B. Limitations

Some limitations come from the system's dependence on the Microsoft Azure cloud service. If the network connection is weak or unstable, there can be noticeable delays in feedback or even occasional failures to process the audio. The Pronunciation Assessment API is mainly trained on adult speech, so its scoring for children may not always match the way a speech therapist would judge the same sounds.

At present, the system only supports English-language assessment, so it is not suitable for children who primarily use other languages. The articulation visuals are currently based on static images, which gives a good idea of how the mouth should look but does not show the full movement over time. Finally, the feedback is fully automated, so it cannot replace the nuanced judgment and emotional support that a human therapist can provide in real sessions. These factors mean that Echo Learn works best as a supplementary tool between therapy visits, not as a full replacement for professional speech support.

X. FUTURE SCOPE

The current design of Echo Learn can be extended in several directions. One possibility is to integrate offline speech-recognition models so that at least basic phoneme-level assessment can continue even when the internet connection is poor or absent. This would also reduce the time between speaking and feedback, making the practice feel more immediate.

The pronunciation models themselves could be fine-tuned using children's speech data, which would likely improve their accuracy for the target user group. The articulation visualization system could be expanded to include simple animations or 3D-style models that show how the tongue moves and how the lips change shape over time, rather than just a single still image. This kind of dynamic feedback would give children a more realistic picture of how sounds are produced.

Supporting more than one language would broaden the platform's usefulness across different communities. The system could also be made more adaptive by using the stored attempt history to select words and sentences that best match the child's current weak sounds, instead of following a fixed list. Machine learning–based analytics could try to predict upcoming difficulties and suggest extra practice before errors become strongly established.

Adding small gamification elements—such as points, badges, or simple rewards for completing practice sessions—might help keep children motivated during longer or repeated sessions. Finally, deploying the application to the cloud would allow usage across multiple devices and make it easier to share progress data between home, school, and therapy rooms.

XI. CONCLUSION

This paper introduced **Echo Learn**, a web-based assistive application that combines Augmentative and Alternative Communication (AAC) support with structured pronunciation practice in one platform. The system fills a clear gap in current AAC tools, which often focus only on expression without offering phoneme-level feedback or visual guidance on how the lips, tongue, and airflow should move. Echo Learn gives children a simple symbol-based board to build sentences and express basic needs, supported by clear speech synthesis tuned for Indian-English conditions, while the pronunciation module uses Microsoft Azure Pronunciation Assessment to detect mispronounced phonemes and guide correction.

The articulation visualization system helps children see how each sound should be physically produced, going beyond what they can learn by listening alone. Testing showed that the system can reliably capture speech, compute phoneme-level scores, and provide targeted guidance that leads to measurable improvements over repeated sessions.

REFERENCES

[1] Microsoft Corporation, "Pronunciation Assessment - Azure AI Services," Microsoft Learn, 2024.

[2] S. Dudy, S. Bedrick, M. Asgari, and A. Kain, "Automatic Analysis of Pronunciations for Children with Speech Sound Disorders," Computer Speech & Language, vol. 50, pp. 62-84, Jul. 2018.

[3] S. M. Witt and S. J. Young, "Phone-Level Pronunciation Scoring and Assessment for Interactive Language Learning," Speech Communication, vol. 30, no. 2-3, pp. 95-108, Feb. 2000.

[4] Y. Zhang et al., "speechocean762: An Open-Source Non-Native English Speech Corpus for Pronunciation Assessment," in Proc. Interspeech, 2021, pp. 456-460.

[5] D. Povey et al., "The Kaldi Speech Recognition Toolkit," in IEEE Workshop on Automatic Speech Recognition and Understanding, 2011, pp. 1-4.

[6] A. Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision," in Proc. ICML, 2023, pp. 1-15.

[7] American Speech-Language-Hearing Association (ASHA), "Speech Sound Disorders: Articulation and Phonology," ASHA.org, 2024.

[8] D. R. Beukelman and P. Mirenda, Augmentative and Alternative Communication: Supporting Children and Adults with Complex Communication Needs, 4th ed. Baltimore, MD: Paul H. Brookes Publishing, 2013

[9] J. Light and D. McNaughton, "Communicative Competence for Individuals Who Require Augmentative and Alternative Communication," Augmentative and Alternative Communication, vol. 30, no. 1, pp. 1-18, Mar. 2014.

[10] J. A. Gierut, "Treatment Efficacy: Functional Phonological Disorders in Children," Journal of Speech, Language, and Hearing Research, vol. 41, pp. S85-S100, Feb. 1998.

[11] H. Franco et al., "The SRI TOPIX System: Automatic Pronunciation Assessment for Children," in Proc. Interspeech, 2010, pp. 1234-1237.