International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I7P110

Volume 74 Issue 7, 146-159, July 2026

Original Article

iRead: A Reading Enhancement Platform with Integrated Small-Vocabulary Speech Recognition for English, Filipino, and Hiligaynon

Jan Carlo T. Arroyo[1*], Bon Eric A. Besonia[2], Allemar Jhone P. Delima[1], Felipe P. Vista IV[1], Mark Ronar G. Galagala[1], Marieth Flor M. Bernardez[2], Shiela Mae H. Espora[2], April Rose A. Zaragosa[1], Rizzamila R. Superio[2]

[1]*College of Information and Computing Studies, Northern Iloilo State University, Estancia, Iloilo, Philippines.*
[2]*College of Education, Northern Iloilo State University, Estancia, Iloilo, Philippines.*

[*]*Corresponding Author : jancarloarroyo@nisu.edu.ph*

Received: 25 February 2026 Revised: 06 June 2026 Accepted: 18 June 2026 Published: 28 July 2026

Abstract - *Reading proficiency is considered a critical educational challenge in a highly multilingual nation such as the Philippines. Digital literacy tools available on the market and those that are found in the literature are mostly English-centric and often lack interactive mechanisms. This study shows the design, technical validation, and implementation of the iRead mobile application software. It is a multilingual mobile reading platform with offline speech recognition function available for three languages, specifically English, Filipino, and Hiligaynon. The mobile application was developed specifically for the Android Operating System using the Flutter framework, while the Vosk API was used for the speech recognition engine. Publicly available pretrained speech recognition models were utilized for English and Filipino languages, while a novel baseline small-vocabulary speech recognition model for Hiligaynon was developed and trained from scratch. A Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) pipeline within the Kaldi framework was then used to form the Hiligaynon speech recognition model. Recognition vocabulary was limited to a 380-word phonics-based lexicon that is aligned with early literacy instruction. Cross-speaker generalization for Hiligaynon was evaluated using a leave-one-speaker-out cross-validation technique across four speakers. Recognition stability was further assessed using standard deviation and confidence interval analysis. The overall system evaluation was conducted using 540 utterances across the three languages under controlled conditions. Recognition performance achieved average accuracies of 92.8% for English, 88.3% for Filipino, and 85.6% for Hiligaynon. Category-level analysis demonstrated the highest performance for vowels, followed by consonants, then consonant–vowel blends. Results suggest that a classical small-vocabulary acoustic model combined with grammar-constrained decoding is technically viable and deployment-ready in a multilingual offline speech-supported literacy app for low-resource educational settings.*

Keywords - *Acoustic Modeling, Grammar-Constrained Decoding, Low-Resource Languages, Mobile Learning, Multilingual Literacy, Offline Speech Recognition.*

1. Introduction

Reading is widely recognized as a foundational skill that underpins learning across all levels of education [1, 2]. It enables learners to access information, construct meaning, and develop critical thinking abilities essential for academic success and lifelong learning [3, 4]. Reading proficiency in early age is critical since difficulties encountered at later educational stages are difficult to correct, which may negatively influence learners' confidence, motivation, and overall academic performance [5]. In the Philippines, reading proficiency among learners remains critically low [6], as evident in national and international assessments conducted. It shows that most Filipino learners fail to meet minimum reading proficiency levels and that many learners are struggling to identify basic information, comprehend explicit meaning, and engage with texts at a functional level [7]. Some interventions were made to increase the reading skills of children, but the result remained unchanged, hence current methods are not optimal to address the same problems [8]. Reading is not a one-size-fits-all learning process, especially in the early years. Problems like difficulty in sound recognition, comprehension, or even decoding words are most common to children sometimes due to the limited opportunities to practice reading using familiar linguistic structure like having those reading materials specific for Filipino and mother tongue languages [9]. Language is critical for reading development in a multilingual context [10, 11]. Children who use reading materials in a different language apart from the dialect used at home present cognitive and affective learning problems. In the Philippines, the use of English as a medium of instruction in schools is prevalent, while Filipino and other regional languages are being used in daily communication.

This leads to having learners exposed to instructional materials that are different from the language they commonly use at home and within their communities, resulting to difficulty in grasping what they read, lose interest in lessons, and taking a longer time to become fluent readers. These affects not only the learning of the children but also their confidence and attitude toward reading and learning [9].

Some public elementary schools in geographically isolated communities in the Philippines has limited instructional resources in regional languages. Teachers improvise on having reading materials since resources are unavailable, which is challenging in multilingual classrooms since learners possess varying levels of exposure to English, Filipino, and their mother tongue languages. Therefore, reading support technologies that accommodate multiple local languages remain highly necessary in early literacy instruction.

With the advent of digital technologies in education, there are now new avenues for addressing long-standing reading literacy challenges [12-14]. Mobile applications offer a flexible learning environment that complements classroom instruction [15, 16]. However, studies show that digital learning tools are still not being optimally used because many institutions lack equipment, internet connection, and devices. Moreover, other teachers and parents are not fully equipped and trained to use these tools effectively [17].

As a result, technology-based learning remains out of reach for many learners, especially in remote or underserved communities in the Philippines [9]. Therefore, there is a need for educational reading tools that are simple, affordable, and do not rely on the internet. Solutions that are low-cost, easy to set up and navigate, and can be utilized offline are preferred. The digital tools give children the opportunity to develop their reading skills, as the tools are more practical for schools and communities with limited resources [18]. However, available digital reading applications still fail to address the current contextual and linguistic challenges.

There are a number of developed mobile applications to improve reading skills, but most of which are focused on English. The available tools are static in nature and do not adequately reflect the local culture, lived experiences, and linguistic backgrounds of Filipino learners [19]. While other digital reading applications provide interactive pronunciation support, only a few are designed for multilingual contexts involving Philippine regional languages. Fewer literature integrates offline speech function for languages such as Hiligaynon, where publicly available language models and annotated dictionaries remain extremely limited. Most of the developed applications use generic English speech recognition rather than phonics-based reading texts for early literacy learning. Thus, many learners in some geographically isolated communities still lack access to culturally relevant and accessible reading technologies. Some available applications also provide limited interactive feedback mechanisms, reducing their effectiveness in supporting reading and pronunciation practice among young learners [9]. Children may experience a disconnect from the learning process if linguistic and cultural backgrounds are not considered, ultimately reducing learning effectiveness and weakening reading motivation [20].

To address the limitations, there is a need to develop educational technologies that are linguistically and culturally aligned with learners in the community they intend to support. Culturally responsive teaching is denoted by the use of materials and lessons that reflect who the learners are, what language they speak, and their experiences in their daily lives [26, 27].

Children have a grasp of the things they learn once they are given stories and texts in their own language and culture, since the content connects to what they already know [28, 29]. Studies show that the use of mother tongue and multiple languages in communities where families face economic challenges and where many different languages are used helps improve reading skills in the early years [8, 9]. Further, interactive feedback is increasingly recognized as a critical component of digital literacy support systems [30]. Implementing interactive mechanisms requires reliable speech recognition functionality that can operate even under constrained infrastructural conditions, such as loss of internet connectivity.

Developing speech recognition tools for Philippine languages like Hiligaynon is difficult because there are no ready-made speech recognition models available for these languages. Further, only a few recorded and labeled speech datasets are available that can be used to train such systems. Furthermore, schools lack the infrastructure needed to run complex speech recognition technology, limiting the development of reading materials that children can use. This limits opportunities for young Filipino learners to practice speaking and reading in their own language.

To address these challenges, this study introduces iRead, a multilingual mobile reading application designed for English, Filipino, and Hiligaynon literacy support. The developed mobile application software uses pretrained speech recognition models for English and Filipino languages and implements a novel and independently trained small-vocabulary model for Hiligaynon. The model was developed from supervised child speech data using a classical Gaussian

Mixture Model-Hidden Markov Model (GMM-HMM) combined with grammar-constrained decoding. In iRead, the speech recognition is limited to a 380-word phonics-based vocabulary only in order to reduce search complexity and improve robustness under small-vocabulary conditions. The architecture also operates fully offline to ensure that the system can be deployed and utilized in areas with limited internet connectivity.

The novelty of this work lies in the integration of multilingual offline speech recognition within a phonics-based mobile reading platform specifically designed for low-resource Philippine language settings.

Unlike many existing literacy applications that rely on internet-based recognition services or support only dominant languages, iREAD integrates pretrained English and Filipino speech recognition models with a newly developed Hiligaynon acoustic model trained using supervised child speech data. The study further demonstrates the feasibility of grammar-constrained decoding for early literacy activities through a limited-vocabulary approach that can be deployed on low-cost Android devices without requiring internet connectivity.

Table 1 presents a summary of related Philippine-based mobile reading applications identified in the literature. Most studies found in the literature focused on English reading applications with limited support for localized languages. On the other hand, this study expands multilingual support by integrating 3 languages commonly used in the Philippines: English, Filipino, and Hiligaynon. The mobile app is further enhanced by dynamic content customization, inclusion of culturally relevant learning materials, and offline functionality.

Table 1. Summary of relevant philippine-based reading mobile applications

Study	Supported Language(s)	Speech Recognition	Content Customization	Culturally-Relevant Topics	Offline Capability
[21]	English	X	X	X	/
[22]	English	X	X	X	/
[23]	English, Filipino	X	X	/	/
[24]	English, Filipino, Bisaya/Cebuano	X	X	X	/
[25]	English, Filipino, Hiligaynon	/	/	/	X
This study	English, Filipino, Hiligaynon	/	/	/	/

In general, this study aimed to develop and evaluate a multilingual mobile reading platform with integrated offline speech recognition for English, Filipino, and Hiligaynon. Specifically, this study aimed to:

1. design and develop a multilingual reading mobile application for early literacy support;
2. develop a small-vocabulary Hiligaynon speech recognition model using supervised child speech data;
3. determine the performance accuracy of the integrated speech recognition modules across vowel, consonant, and consonant–vowel blend activities; and
4. Evaluate the cross-speaker generalization capability of the Hiligaynon speech recognition model using a leave-one-speaker-out validation approach.

The remainder of this paper is organized as follows: Section 2 describes the system design, vocabulary specification, speech recognition integration, and evaluation process of iRead.

Section 3 presents the speech recognition performance results and analysis. Finally, Section 4 concludes the study and outlines directions for future work.

2. Methods

2.1. System Design and Implementation

This study followed a developmental research approach in the design, implementation, and technical evaluation of a multilingual reading mobile application that integrates an offline speech recognition. The development process of iRead followed the major phases of the System Development Life Cycle (SDLC). The processes involved requirements analysis, system design, implementation, and testing. Tests were conducted to assess the functionality and recognition accuracy of the speech-enabled modules under controlled conditions.

The Flutter framework was used to develop the mobile application for it to be compatible and deployable on the Android Platform. The offline speech recognition capability of the mobile application was implemented by using the Vosk API integrated with Kaldi-based acoustic models. For the speech recognition, publicly available pretrained models were utilized for detecting English and Filipino words. Then, a custom small-vocabulary speech recognition model was independently developed for detecting Hiligaynon words. The system architecture of the iRead mobile application consists of 3 primary modules. The first module is the language selection module. The second module is the reading content module, and the third is the speech recognition module. The functional testing in this study focused on pronunciation-based phonics activities involving vowels, consonants, and consonant–vowel blends. This study did not involve any classroom intervention or longitudinal learning assessment yet; rather, the investigation primarily focused on technical performance, offline deployment capability, and multilingual speech recognition accuracy.

2.1.1. System Architecture Overview

The overall system architecture of the iRead mobile application uses a client-side offline processing framework. Stored locally within the mobile device are the reading materials, pronunciation prompts, vocabulary data, and speech recognition resources. This approach minimizes the dependence on internet connectivity. During speech interactions, the audio captured through the device microphone undergoes feature extraction and decoding locally using the integrated Vosk speech recognition engine. For the English and Filipino speech recognition tasks, the pretrained models were integrated directly into the application environment. For Hiligaynon, a dedicated acoustic model was developed and converted into a Vosk-compatible runtime format after a Kaldi-based training. Grammar-constrained decoding was implemented to reduce recognition ambiguity and improve stability during the phonics exercises involving isolated word utterances. The offline-capable architecture was intentionally selected to support the deployment of the mobile application in low-resource educational settings. This is important for schools where the internet connectivity may be unstable, intermittent, or even unavailable.

2.2. Target Users of the System

The target of the iRead mobile application was designed for 3 primary user groups, to wit: learners, parents, and teachers. The primary end-users are the learners who interact directly with the reading content, language selection features, and speech-enabled pronunciation support modules.

The system interface of iRead was designed to be visually clear, intuitive, and appropriate for early-age users. Another target user are the parents who are considered as the key facilitators of learning, particularly in home-based reading environments. The mobile application was designed to allow the parents to guide and support their children during reading activities without requiring any advanced technical expertise. Lastly, the third target users are the teachers who may use the mobile application as a supplemental tool for reading in the classroom since it offers multilingual and culturally relevant content that supports learning. The iRead mobile application was primarily intended to be used by beginning or struggling readers, where foundational literacy instruction should be emphasized. The design of the mobile application user interfaces considered age-appropriate interaction principles such as simplified menus, large touch elements, minimal text clutter, and multilingual accessibility. The parents and teachers were included as users to support guided reading activities in both home and classroom settings.

2.3. Phonics Vocabulary Specification

The Hiligaynon speech recognition feature integrated in the mobile application was built using a set of recorded phonics-based words that support early reading instruction. The vocabulary included 5 vowels, 15 consonants, and 75 consonant–vowel blends. For each phonics character, 4 sample words were identified based on how clear they sound, how consistently they are spelled, and how familiar they are to young Filipino learners. These words were carefully selected to reflect commonly used and culturally familiar terms found in local early reading materials. Vocabulary selection was guided by early literacy principles emphasizing phonetic consistency, familiarity, and ease of pronunciation among young learners. Priority was given to commonly encountered words that reflect everyday linguistic exposure within Filipino and Hiligaynon-speaking communities.

Limiting the vocabulary to a total of 380 words, this helped reduce decoding complexity and improve recognition stability under constrained conditions. All vocabulary items were saved in UTF-8 format and stored locally using structured JSON files. Table 2 shows the distribution of the vocabulary.

Table 2. Hiligaynon phonics vocabulary distribution

Phonics Category	Number of Units	Words per Unit	Total Words
Vowels	5	4	20
Consonants	15	4	60
Consonant–Vowel Blends	75	4	300
Total	95	—	380

2.4. Baseline Hiligaynon Speech Recognition Model Development

In order to build a speech recognition system for the Hiligaynon language, a recording of 380 Hiligaynon words was performed. The generated small-vocabulary model was developed using the Kaldi speech recognition toolkit. After training, the model was converted and integrated into the Vosk runtime environment, allowing it to run fully offline within the iRead mobile application. Unlike the English and Filipino languages, which already had existing publicly available pretrained models available over the internet, Hiligaynon did not have such resources. Hence, there was a need to create and train the Hiligaynon speech recognition model from scratch using a supervised learning approach using the recorded speech data.

The decision to utilize a classical GMM–HMM framework in this study instead of using a deep neural network architecture was primarily because of the dataset size limitations. Given the relatively small number of training utterances available, the selected approach offered a better training stability, a lower computational requirement, and a reduced risk of overfitting. Previous studies involving low-resource speech recognition also have similarly demonstrated that classical acoustic modeling approaches remain viable under constrained data conditions.

2.4.1. Speech Corpus Collection and Preparation

In order to train the Hiligaynon speech recognition model, a supervised speech dataset was created and collected from 4 child participants. These participants were recruited for supervised speech recording activities. The participants consisted of 2 male and 2 female child speakers who were native or familiar speakers of Hiligaynon. Participation was voluntary and conducted with parental guidance and consent prior to the recording. To protect participant privacy, no personally identifying information was stored within the speech dataset. The audio recordings were collected solely for research and acoustic modeling purposes. Each speaker produced 1 utterance for each of the 380 predefined words. After the recording, a total of 1,520 utterances (380 words × 4 speakers) formed the dataset.

All of the voice recordings were performed and captured in a quiet room to avoid any background noise and echoes. During the recording, a microphone was placed about 20cm to 30cm away from the speaker's mouth. The audio was recorded at 16,000 Hz using a 16-bit PCM mono format, which is considered the standard setting for small-vocabulary training using the Kaldi toolkit. Each of the word recordings lasted between approximately 1 to 3 seconds. Each audio file was checked manually and matched with its written word to ensure labels were correctly annotated. The details of the training data are shown in Table 3.

Table 3. Hiligaynon acoustic training corpus characteristics

Parameter	Value
Number of Speakers	4
Gender Distribution	2 Male, 2 Female
Vocabulary Size	380 Words
Utterances per Word	1
Total Utterances	1,520
Utterance Duration	1.5–3.0 seconds
Sampling Rate	16,000 Hz
Audio Format	16-bit PCM, Mono
Recording Environment	Quiet Indoor

Since the dataset is rather small, a classical GMM–HMM approach was used for the training. This method helped reduce the number of model parameters, lowered the risk of overfitting, and kept the model stable even with limited training data.

Ethical Considerations

Ethical considerations were carefully observed throughout the collection of speech data and the system evaluation process. Prior to participation, informed consent was secured from the parents or guardians of all children participants involved in the recording activities. Participants were informed regarding the purpose of the recordings, the intended research use of the collected audio data, and the voluntary nature of participation. The collected recordings were used exclusively for acoustic model development and evaluation. No sensitive personal identifiers were included in the dataset, and all recordings were stored in secure local research storage accessible only to the researchers. The study focused solely on isolated-word pronunciation tasks and did not involve behavioral, psychological, or medical assessment of participants.

2.4.2. Speech Recognition Modeling Pipeline

The speech recognition model for Hiligaynon was built using GMM–HMM within the Kaldi toolkit because of its ability to work even with a small vocabulary dataset. A neural network was not used because the dataset was too small to support deep learning methods.

For each of the recordings, 13 Mel-Frequency Cepstral Coefficients or MFCCs were extracted. Delta and delta-delta coefficients were added to make a total of 39 features per frame. A 25 millisecond window was used with a 10 millisecond frame shift. Cepstral Mean Normalization was applied to reduce differences caused by the recording environment.

The Kaldi toolkit's standard setting was used in training the Hiligaynon speech recognition model. The training sequence progressed from monophone models to context-dependent triphone models. Linear Discriminant Analysis (LDA) and Maximum Likelihood Linear Transform (MLLT) were then applied for feature transformations. Speaker Adaptive Training (SAT) was performed using feature-space Maximum Likelihood Linear Regression (fMLLR) to reduce inter-speaker variability and improve generalization.

At each step, alignments were recomputed before updating the model. In each cross-validation fold, steps like normalization, clustering, and speaker adaptation were done using only the training data. This ensured that no information from the test data leaked into the model. The default settings for Kaldi were used throughout the training without any manual tuning.

A pronunciation lexicon was constructed directly from the predefined 380-word vocabulary. Phoneme mappings were defined according to the documented Hiligaynon phonemic inventory to ensure consistent grapheme-to-phoneme alignment. After training, the final speech recognition model was compiled into a format compatible with the Vosk API to allow the speech recognition module of the iRead to be accessible by the users even without the internet.

2.4.3. Grammar-Constrained Decoding Configuration

To manage search complexity and improve recognition accuracy, decoding was carried out using grammar-constrained search. Open-vocabulary decoding was turned off. For each pronunciation task, the system limited possible word matches to the full 380-word vocabulary. When a learner was prompted to pronounce a word, the decoding pattern used is: <sil> target_word <sil>, where <sil> means silence or a pause. This approach reduced the number of possible word choices and made recognition more stable, given the small vocabulary size.

2.5. Cross-Speaker Evaluation Procedure

To test the performance of the Hiligaynon model and to prevent overfitting, a Leave-One-Speaker-Out (LOSO) cross-validation technique was performed. Four experimental folds were conducted since there were four speakers in the dataset. In each fold, acoustic model training was performed using data from 3 speakers only, for a total of 1,140 utterances. The remaining speaker dataset, comprising 380 utterances, was used for testing. This process was repeated until each speaker dataset had been the test subject.

Recognition performance was measured by looking at whether each word was correctly recognized. Since each of the test utterances consisted of a single word, Word Error Rate (WER) corresponded directly to word-level accuracy. No other evaluation methods, like confidence thresholding, phoneme-level scoring, or partial credit, were applied. A word was counted as correct only if it matched the target exactly.

For each fold, overall accuracy was computed across the test set. Final performance was reported as the mean accuracy across all four folds. To further assess model consistency, standard deviation values were computed across validation folds. Recognition accuracy was selected as the primary evaluation metric because each utterance corresponded to a single isolated vocabulary item under grammar-constrained decoding conditions. In every round, all model training steps were performed using only the three training speaker datasets. The test set was not used at any point during training. This kept the test results clean and independent. Table 4 shows the full evaluation setup. Further, LOSO validation was performed exclusively for Hiligaynon because English and Filipino utilized pretrained speech recognition models whose original training corpora were not accessible.

Table 4. Leave-One-Speaker-Out evaluation configuration

Parameter	Value
Total Speakers	4
Training Speakers per Fold	3
Training Utterances per Fold	1,140
Test Speaker per Fold	1
Test Utterances per Fold	380
Validation Strategy	Leave-One-Speaker-Out
Evaluation Metric	Utterance-Level Word Accuracy
Decoding Mode	Grammar-Constrained
Deployment Mode	Offline

2.6. Statistical Treatment of Data

Descriptive statistical methods were used to summarize recognition performance across languages and phonics categories. Accuracy percentages were computed by dividing the number of correctly recognized utterances by the total number of prompts evaluated. Mean recognition accuracy and standard deviation values were calculated for the leave-one-speaker-out validation experiments. Comparative analysis across phonics categories was performed descriptively to examine recognition behavior under varying speech complexity conditions. Additional statistical descriptors, including standard deviation, confidence intervals, and fold variability measures, were incorporated to further assess model consistency and recognition stability.

3. Results and Discussion

3.1. Speech Recognition Performance Across Categories

In order to assess the performance of the English, Filipino, and Hiligaynon speech recognition models, a series of structured repetition tasks was performed. These tests involved 4 children to utter 45 prompts per language. Each child uttered 10, 15, and 20 prompts for vowels, consonants, and consonant–vowel blends, respectively. A total of 40 vowels, 60 consonants, and 80 blend utterances per language were obtained. This provided a total of 180 utterances for English, Filipino, and Hiligaynon. The speech recognition accuracy for vowel prompts across all categories is presented in Table 5. As observed, the model with the highest vowel recognition accuracy is English at 97.5%, followed by Filipino at 92.5% and Hiligaynon at 90.0%. Because vowel sounds have stable acoustic features and were produced in fairly consistent ways by the speakers, they were found to be generally easier for the model to recognize.

Table 5. Speech recognition accuracy for vowel prompts

Language	Total Utterances	Correctly Recognized	Accuracy (%)
English	40	39	97.5
Filipino	40	37	92.5
Hiligaynon	40	36	90.0

As for the tests involving consonants, speech recognition models found this task to be more complex than vowels, given the lower accuracies obtained. English consonants obtained the highest accuracy at 93.3%, while Filipino achieved 88.3%. Hiligaynon obtained the least accuracy at 85.0%, which is slightly lower than Filipino. This result makes sense because the consonant sounds require more precise mouth movements and vary more in how they are voiced and where they are produced. Nevertheless, looking at the results, all of the 3 models remained above 85%. This means that the system works well enough for simple phonics practice where the children repeat sound prompts. Table 6 shows the complete results for consonant recognition. In terms of recognizing consonant-vowel blends, performance accuracy across all 3 languages even dipped compared to vowels and consonants. The English blends obtained the highest accuracy of 90.0%, while Filipino and Hiligaynon only obtained 86.3% and 83.8%, respectively. Unlike the vowels or consonants, blends involve shifting sounds and greater timing differences, which make them more difficult for speech recognition systems to process. Table 7 shows the complete results for the consonant-vowel blends.

Table 6. Speech recognition accuracy for consonant prompts

Language	No. of Prompts	Correctly Recognized	Accuracy (%)
English	60	56	93.3
Filipino	60	53	88.3
Hiligaynon	60	51	85.0

Figure 1 illustrates the speech recognition accuracy obtained across all of the phonics categories for English, Filipino, and Hiligaynon. Results show that vowel recognition obtained the highest accuracies across all languages, while consonant–vowel blends produced comparatively lower performance due to their increased phonetic complexity.

Table 7. Speech recognition accuracy for consonant–vowel blend prompts

Language	No. of Prompts	Correctly Recognized	Accuracy (%)
English	80	72	90.0
Filipino	80	69	86.3
Hiligaynon	80	67	83.8

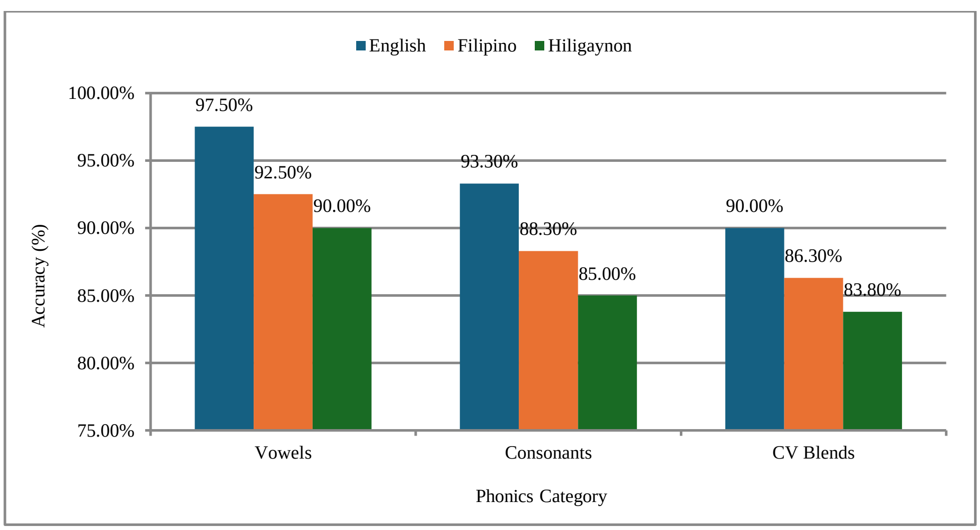

Fig. 1 Speech recognition accuracy by phonics category

3.2. Overall Recognition Performance

Based on the test results, it can be observed that accuracy decreased as sounds became more complex, regardless of the language being tested. The differences across phonics categories also indicate practical performance effects associated with phonetic complexity. Recognition accuracy consistently declined from the vowels to the consonants and further decreased for the consonant–vowel blends in all the 3 languages. This pattern suggests that articulation has an influence on the recognition behavior of the language model. Table 8 and Figure 2 show the summary of recognition results across all the categories. The speech recognition performance for the English model obtained the highest recognition accuracy at 92.8%. This is followed by 88.3% for Filipino and 85.6% for Hiligaynon. Even with the limited availability of the dataset, the Hiligaynon model obtained a decent overall accuracy above 85% and above 80% in each category. This shows that implementing a classic GMM–HMM approach for small vocabulary speech recognition tasks is feasible. Since the vocabulary used in this study was limited to 380 words, this helped reduce confusion and made the speech recognition more stable, especially on mobile devices without an internet connection. The 95% confidence intervals were computed for each language model. The English model produced the narrowest confidence interval, indicating the most stable recognition performance across the evaluated prompts. Slightly wider intervals were observed for Filipino and Hiligaynon, which may reflect increased pronunciation variability and the smaller locally collected training dataset used for the Hiligaynon acoustic model.

Table 8. Overall speech recognition accuracy

Language	Total Prompts	Correctly Recognized	Mean Accuracy (%)	95% CI
English	180	167	92.8	±3.8
Filipino	180	159	88.3	±4.7
Hiligaynon	180	154	85.6	±5.1

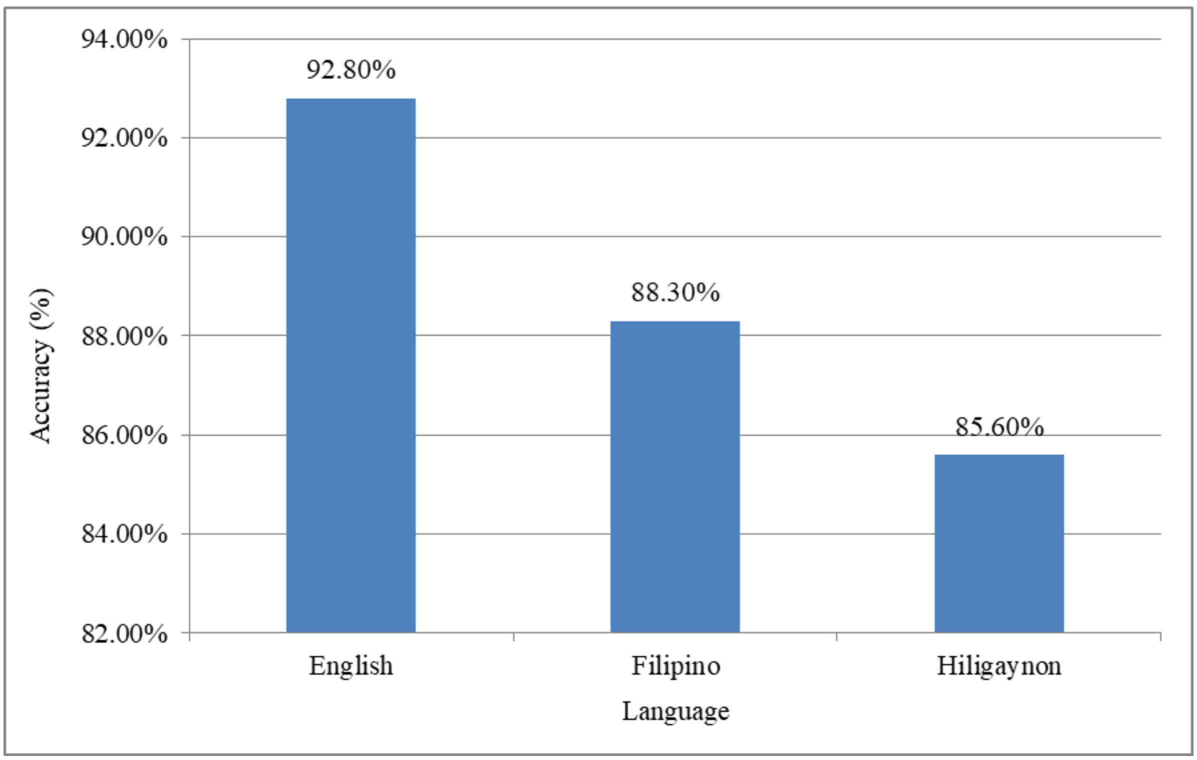

Fig. 2 Overall speech recognition accuracy

After performing all of the tests, a pattern was revealed. Vowels were easiest, consonants were moderate, and consonant-vowel blends were hardest to recognize. This suggests that how complex a sound is matters more than the language it comes from or the model's architecture. Child speech naturally varies, and that variation affects recognition most in the more complex sounds.

Overall, the results of the various tests show that a small-vocabulary speech recognition model can deliver a reliable performance for reading enhancement even in low-resource languages like Hiligaynon. Implementing grammar-constrained decoding makes it even more reliable by limiting what the system needs to guess, which is perfect for an offline mobile learning application.

3.3. Leave-One-Speaker-Out Cross-Validation Results

Results of the LOSO cross-validation tests revealed an average accuracy of 85.6% with a standard deviation of only 0.55. The corresponding co-efficient of variation was below 1%, indicating low fold-to-fold variability and suggesting relatively stable recognition behavior across different speaker partitions.

This means that the model's performance was very consistent across different speakers, given the small dataset used for training. Table 9 and Figure 3 show these results. The 0.55 standard deviation shows that the speech recognition model behaved consistently despite the changes in test speaker selection. This means that the speech recognition model captured phonetic information reasonably well across the participating child speakers. The limited number of speakers remains a constraint. A broader demographic variation may influence performance differently in real-world deployment settings. Because the speech corpus was collected from a limited regional speaker group under controlled conditions, the resulting model may exhibit performance bias when applied to speakers with substantially different accents, speech pacing, or in varying environmental conditions.

Table 9. LOSO Fold-Level recognition accuracy for hiligaynon

Fold	Held-Out Speaker	Test Utterances	Correct	Accuracy (%)
1	Speaker A	380	324	85.3
2	Speaker B	380	328	86.3
3	Speaker C	380	323	85.0
4	Speaker D	380	326	85.8
Mean	—	—	—	85.6
SD	—	—	—	0.55

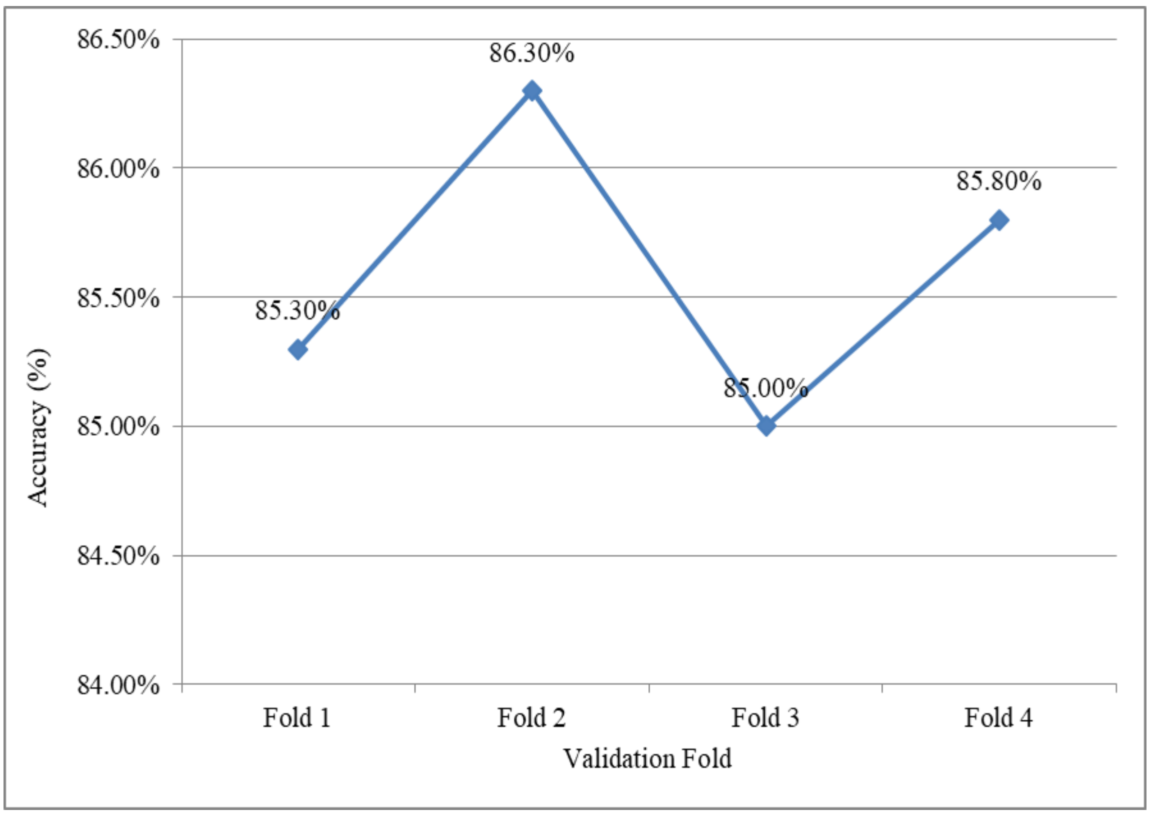

Fig. 3 LOSO Fold-Level recognition accuracy for hiligaynon

3.4. Application Launch and Splash Screen Design

Upon opening the mobile application, a splash screen with the iRead logo greets the user, as shown in Figure 4. The app loaded all necessary resources and settings without delays or errors, then moved smoothly to the main screen. All language options, whether in English, Filipino, and Hiligaynon, were given equal importance and visual treatment. No language stood out more than the others. Cultural details, such as colors, patterns, images, and overall look, were thoughtfully chosen to feel warm and community-oriented.

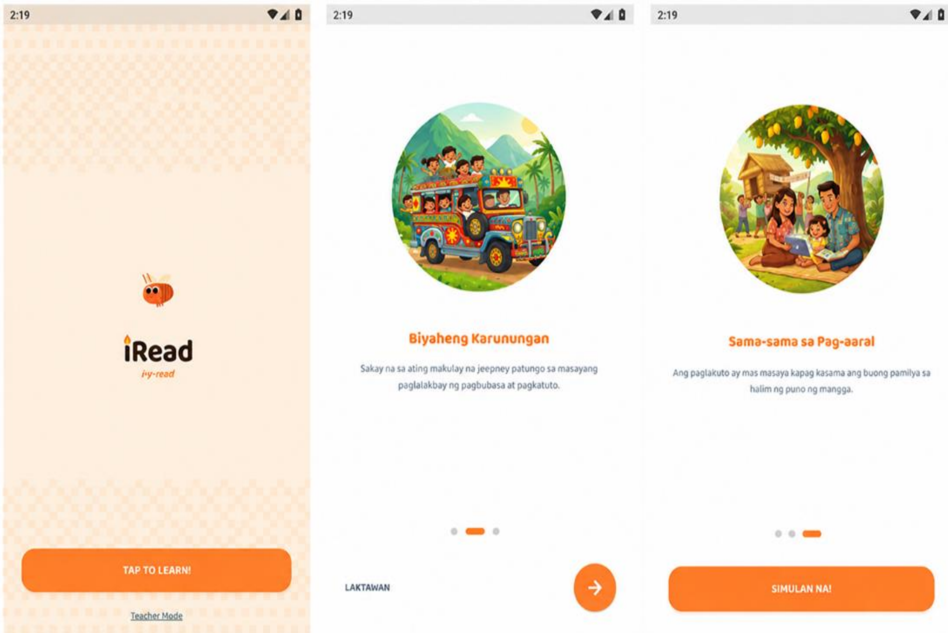

Fig. 4 iRead splash screens

3.5. Language Selection Interface

In order for users to practice their reading skills, they must first select the language they want to learn, as shown in Figures 5 and 6. Functional testing confirmed that this feature works as expected. There is no need to close and reopen the app when a learner selects and confirms a language of their choice. The contents on the succeeding screens update immediately and smoothly based on the choice made.

Ease of use was the primary principle followed when designing the language selection screen. The buttons are large enough with clear labels, and there is limited text only.

The navigational menu is simple for young learners to use on their own, while still being easy for parents and teachers to understand when helping their children.

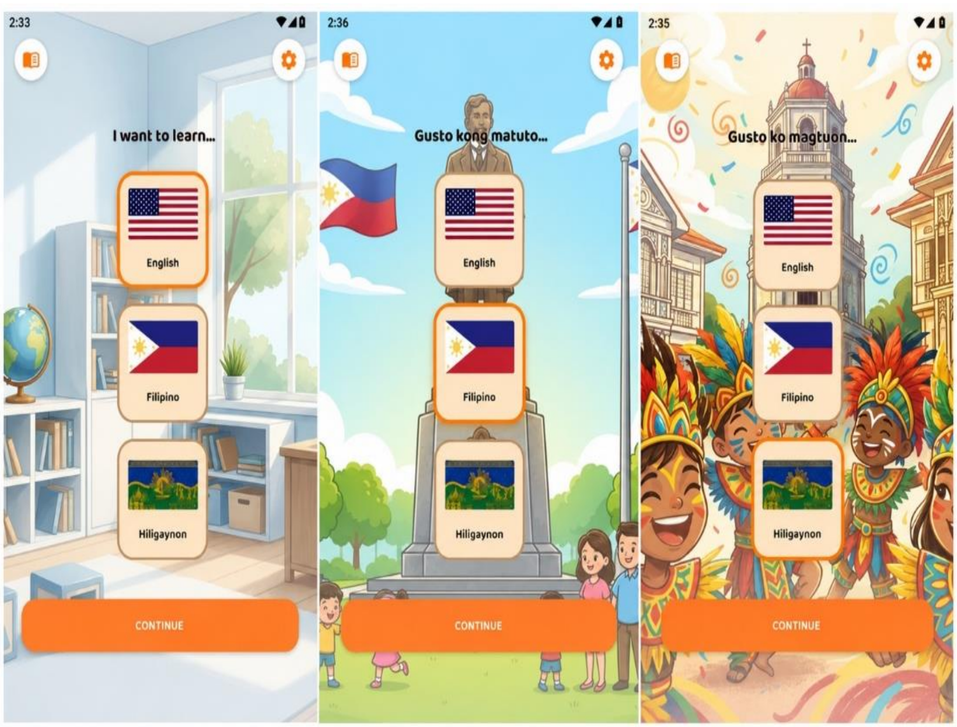

Fig. 5 Language selection interfaces

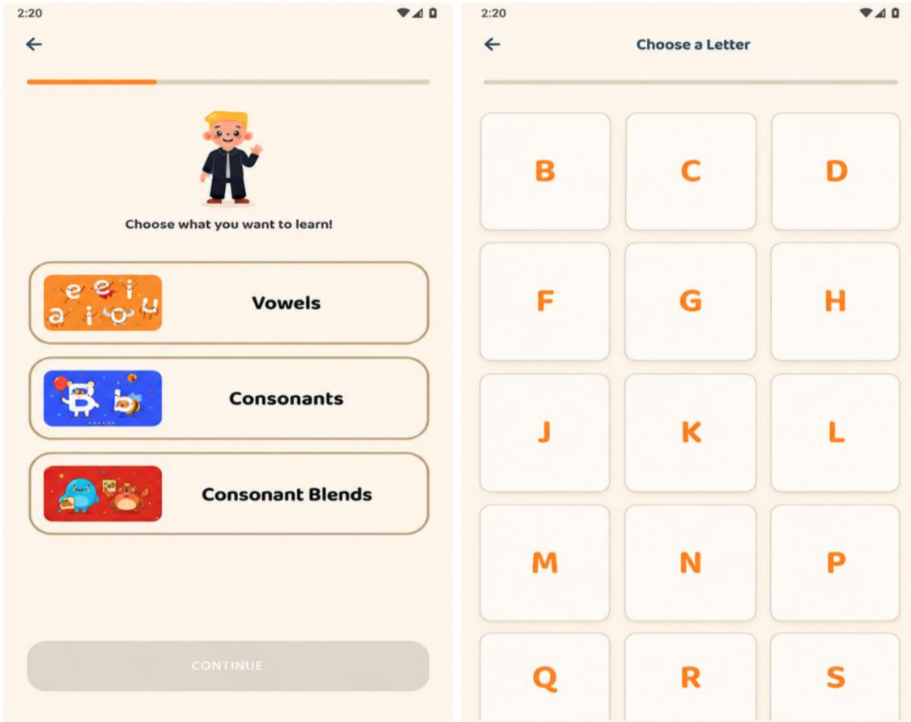

Fig. 6 Word-Consonant-Blend selection interfaces

3.6. Speech Recognition and Oral Reading Interaction

One of the most significant contributions of this study is the successful integration of an offline speech recognition module for English, Filipino, and Hiligaynon in one system. The mobile application has a feature that captures spoken words from a device's microphone and then converts them into text during reading activities. Feedback is also provided to evaluate whether the pronunciation uttered is correct or not. Figure 7 shows an example of how the mobile application implements this interactive feature.

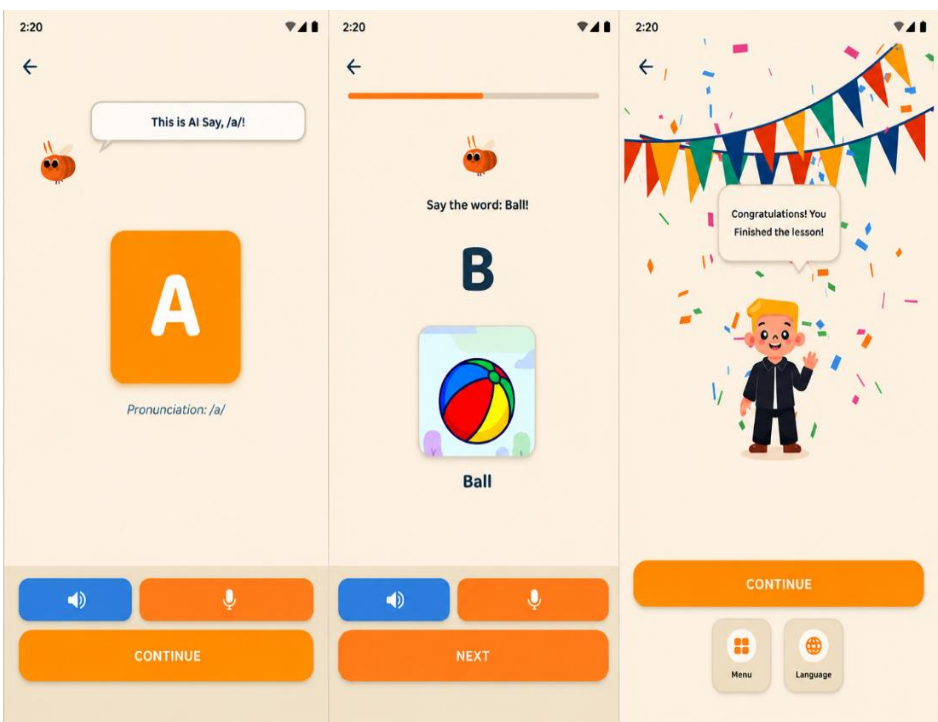

Fig. 7 Speech recognition interfaces

The speech recognition feature of the mobile application is working without any problems encountered. This means that it is possible to add speech recognition for local languages despite a lack of database resources.

Since the mobile application can be accessible even offline, it may be deployed in places where the internet service is slow or even limited. This makes iRead a practical educational tool for reading in many remote or rural schools and communities in the Philippines.

3.7. Reading Content

The iRead mobile application is also like a mini reading library, which has various reading materials for young learners. The mobile application has a variety of local legends, folklore, and everyday themes that reflect life in Philippine communities. These resources were thoughtfully added since these contents are familiar to Filipino learners. The stories also feature an automatic translation mechanism so that the users can read the same story in either English, Filipino, or Hiligaynon, as shown in Figure 8.

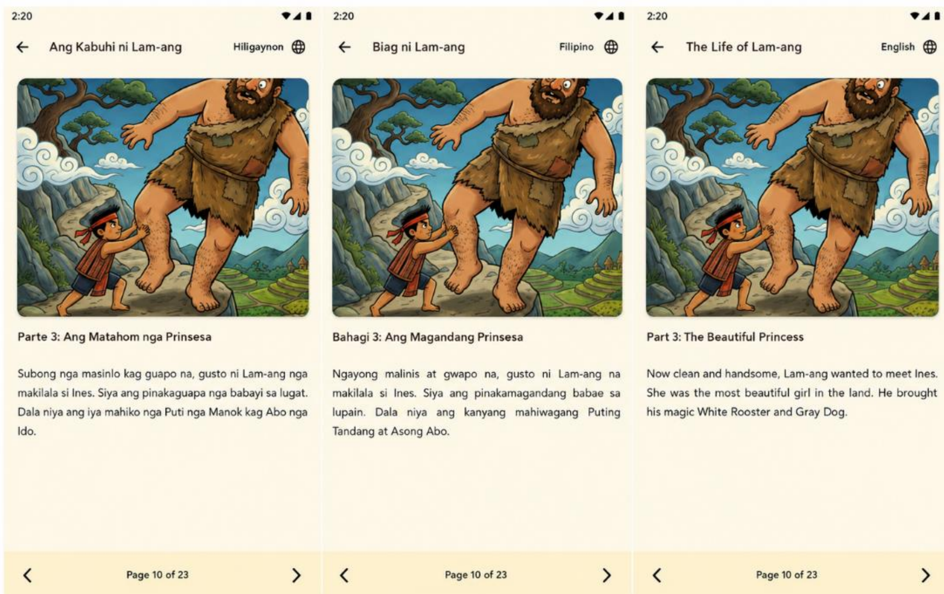

Fig. 8 Reading materials interfaces

The use of text, images, and content layouts was carefully considered to align with the interests of young children. The results of the functional test show no technical difficulties experienced in navigating between the various storybooks.

Browsing the mini library and selecting the stories to read were also smooth, and no glitches or delays were experienced. Since the mobile application is multilingual, stories are available in English, Filipino, and Hiligaynon, which users can easily switch to between versions based on their preferences. This feature allows learners, parents, and teachers to easily access stories that are familiar and culturally relevant to them.

3.8. Study Limitations

This study did not evaluate continuous speech recognition, spontaneous reading, or noisy classroom environments since it is limited to isolated-word phonics activities under controlled recording conditions. Consequently, the reported recognition accuracy should be interpreted within the scope of constrained pronunciation exercises only.

Another limitation is noted, including the Hiligaynon speech recognition model trained using fewer supervised dataset collected from only four speakers. Although cross-validation procedures were applied to reduce overfitting, different users may produce varying speech recognition behaviors.

Lastly, the results of this study focused only on the technical validation rather than educational effectiveness. While the application demonstrated functional multilingual speech recognition capability, the study did not measure reading comprehension improvement, literacy, or long-term learner outcomes among actual classroom users.

4. Conclusion and Recommendations

In this study, a mobile app for reading enhancement with integrated speech recognition and offline capability was introduced. The app was designed to work in 3 languages: English, Filipino, and Hiligaynon. The Flutter framework was used to develop the app, which was subsequently deployed to Android. The app's speech recognition feature was implemented using the Vosk API. Since there were already publicly available pretrained speech recognition models for

English and Filipino, these models were directly used in the system. However, for Hiligaynon, a new small-vocabulary speech recognition model had to be trained from scratch using a child speech dataset and following a GMM–HMM pipeline via the Kaldi toolkit.

A total of 1,520-word recordings or utterances were used to train the Hiligaynon model. To test the model's efficacy and verify its generalization, a leave-one-speaker-out cross-validation approach was used. Overall, the speech recognition models achieved average accuracies of 92.8% for English, 88.3% for Filipino, and 85.6% for Hiligaynon. In all 3 languages, vowels were found to be easiest to recognize, followed by consonants. Consonant-vowel blends were found to be the hardest, resulting in the lowest accuracies among all categories. This makes sense given how much more complex those sounds are, especially in child speech.

As for the Hiligaynon speech recognition model, its accuracy remained above 85% overall and above 80% in every phonics category, given its limited dataset. This result suggests that the GMM–HMM approach can still work well for small-vocabulary speech recognition training of under-resourced languages. The use of the LOSO validation approach also confirms that the results reflect generalization, not just memorization of the training data. Since the mobile application was developed to be offline compatible, it is well-suited to be deployed in schools and communities where internet access is unreliable or limited. This matters in the Philippines, where connectivity remains a significant issue in many remote or rural communities. It is also important to note that the application was designed for speech recognition tasks with single words only. Including spontaneous speech, continuous reading, or open-vocabulary recognition will require a separate set of training processes with more data and advanced algorithms.

Based on the successful development of iRead, there are several recommendations for future work. First, it is recommended to improve the capability of the speech recognition to handle varying accents, age groups, and speaking styles by incorporating more Hiligaynon training data. Another recommendation is to include features that allow for the processing and recognition of longer speech data, such as phrases, sentences, or even paragraphs. Next is to test the mobile app in real-world environments with actual learners, parents, and teachers to see how well it works in school and home settings and whether people find the app valuable and easy to use. Finally, it is recommended to provide orientation and support to the teachers and the parents to help them use the app more effectively.

Conflicts of Interest
The authors declare no conflict of interest in the publication of this paper.

Funding Statement
The authors express their sincere appreciation to Northern Iloilo State University (NISU) for the institutional support and funding provided in the conduct of this study.

References
[1] Debra D. Murphy, *Teaching Literacy Foundations*, 2nd ed., Arkansas Tech University, 2025. [Google Scholar] [Publisher Link]

[2] Lhoussine Qasserras, “Beyond the Words: The Art and Science of Teaching Reading Interactively,” *European Journal of Education Studies*, vol. 10, no. 12, pp. 182-210, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ambar Valeria Velez Parrales, “*The Effect of Reading Activities in the Development of Critical Thinking in 10th Grade Students from Virgen Del Cisne School*,” Bachelor Thesis, La Libertad: Santa Elena Peninsula State University, pp. 1-64, 2023. [Google Scholar] [Publisher Link]

[4] Desti Angraini et al., “A Reading Teacher’s Perspective in Constructing the Students’ Critical Thinking,” *SALEE: Study of Applied Linguistics and English Education*, vol. 5, no. 1, pp. 290-305, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Alona Medalia Cadiz-Gabejan, and Madelyn C. Quirino, “Students’ Reading Proficiency and Academic Performance,” *International Journal of English Language Studies*, vol. 3, no. 6, pp. 30-40, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Niña Keshe Librea et al., “Low Reading Literacy Skills of Elementary Pupils in the Philippines: Systematic Review,” *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, vol. 11, no. 4, pp. 1978-1985, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Judith Cagaanan, “When Schooling Does Not Translate to Learning: Localizing the World Bank’s Learning Poverty Framework for SDG 4,” *Langkit: Journal of Social Sciences and Humanities*, vol. 14, no. 2, pp. 93-110, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] OECD, “*PISA 2022 Results (Volume I): The State of Learning and Equity in Education*,” Organisation for Economic Co-Operation and Development, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Arthea Arese Quesada et al., “Tech-Mediated Learning Resources for Developing Foundational Reading Literacy Skills in K to 3: Literature Review, Infrastructure Situation, Materials Availability Review, and Policy Recommendations for Philippine Schools,” *The Second Congressional Commission on Education (EDCOM II)*, 2026. [Google Scholar] [Publisher Link]

[10] Christopher J. Wagner, “Reading Identities Across Language Contexts: The Role of Language of Text and Talk for Multilingual Learners,” *Journal of Early Childhood Literacy*, vol. 23, no. 2, pp. 317-343, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Khritish Swargiary, *Language and Learning: The Crucial Role of Language in the Teaching-Learning Process*, Google, 2024. [Google Scholar]

[12] Halyna Kuznetsova et al., "Incorporating Innovative Technologies into Higher Education Teaching: Mastery and Implementation Perspectives for Educators," *Multidisciplinary Reviews*, vol. 7, no. 1, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Christine N. Green, "Strategic Considerations for Integrating Digital Technology and Literacy Learning," *The Language and Literacy Spectrum*, vol. 35, no. 1, pp. 1-37, 2024. [Google Scholar] [Publisher Link]

[14] Mohamad Ahmad Saleem Khasawneh, and Barana Belton, "Disclosing the Effects of Automated Feedback on Reading Comprehension, Reading Motivation, Reading Engagement, and Reading Anxiety Through Personalized Technology-Enhanced Learning," *Computers in Human Behavior Reports*, vol. 20, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] Liza Husnita et al., "The Role of Mobile Technology in Improving Accessibility and Quality of Learning," *Jurnal Manajemen Pendidikan*, vol. 11, no. 2, pp. 259-271, 2023. [Google Scholar]

[16] Mohammad Abduljawad, and Amal Ahmad, "An Analysis of Mobile Learning (M-Learning) in Education," *Multicultural Education*, vol. 9, no. 2, pp. 145-152, 2023. [CrossRef] [Google Scholar]

[17] Al-Rashiff Hamjilani Mastul, Claze Therese de Vera, and Cristina Beltran Jayme, "Understanding the use of EduTech in Schools in the Philippines: Recommendations for Effectiveness," *Engineering Science Letter*, vol. 2, no. 2, pp. 37-40, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Rebecca Ferguson, Leigh-Anne Perryman, and Simon J. Ball, "The Importance of Offline Options for Online Learners," *Journal of Interactive Media in Education*, vol. 2024, no. 1, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Alyssa Arbuckle, Janneke Adema, and Élika Ortega, "Editors' Gloss: The Problem with Monolingualism in Academic Knowledge Production," *The Journal of Electronic Publishing*, vol. 27, no. 1, pp. 33-40, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Jenny Eppard et al., "EdTech Culturation: Integrating a Culturally Relevant Pedagogy into Educational Technology," *Electronic Journal of e-Learning*, vol. 19, no. 6, pp. 516-530, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Mary Jane D. Quimsing, and Ruth Ortega-Dela Cruz, "Technology-based Reading Application on Improving Reading Literacy Level Among Struggling Readers in a Public Elementary School in the Philippines," *Pedagogy: Journal of English Language Teaching*, vol. 12, no. 1, pp. 15-28, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Joseph Arsie C. Cabantog, "I-READ: Interactive Reading Mobile Application for Enhanced Senior High School Reading and Writing Skills," *Journal of Interdisciplinary Perspectives*, vol. 3, no. 10, pp. 6-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Mia Coleen C. Geca et al., "LiteraturaPH: An Adapted System Analysis and Design-based Mobile Application for Improving Reading and Comprehension using Philippine Folk Literature among Primary Education Students," *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Ho Chi Minh City, Vietnam, pp. 183-195, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Caroline Rose C. Marvas et al., "Enhancing Reading Comprehension in Grade 5 Students using the Kiddie ReCom App," *Davao Research Journal*, vol. 15, no. 3, pp. 100-110, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Rosie Jane P. Siosan, Joel T. De Castro, and Y.C. Byun, "Design and Development of a Synergistic Educational Tool with Hiligaynon-English Recurrent Neural Network Model," *Journal of Physics: Conference Series*, vol. 2559, no. 1, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Sydney Snyder, and Diane Staehr Fenner, *Culturally Responsive Teaching for Multilingual Learners: Tools for Equity*, Corwin Press, 2021. [Google Scholar] [Publisher Link]

[27] Yi-Huang Shih, "Culturally Responsive Curriculum: A Systematic Overview," *Edelweiss Applied Science and Technology*, vol. 8, no. 6, pp. 6917-6925, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[28] Chima Abimbola Eden, Oyebola Olusola Ayeni, and Nneamaka Chisom Onyebuchi, "Cross-Cultural Instructional Design: A Framework for Multilingual and Interdisciplinary Education," *International Journal of Frontiers in Science and Technology Research*, vol. 6, no. 2, pp. 60-70, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Nanda Dimitrov, and Aisha Haque, "Intercultural Teaching Competence: A Multi-Disciplinary Model for Instructor Reflection," *Intercultural Education*, vol. 27, no. 5, pp. 437-456, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[30] Kurniawan Arif Maspul, "Using Interactive Language Development Tools to Enhance Cognitive and Literacy Skills in K-12 Education," *J-SHMIC: Journal of English for Academi*, vol. 11, no. 1, pp. 20-35, 2024. [CrossRef] [Google Scholar] [Publisher Link]