Mobile Base Sinhala Book Reader for Visually Impaired Individuals

Project ID: 2023-198

Project Proposal Report

Jayathunga T.M. – IT20146238

Supervisor: Prof. Koliya Pulasinghe

B.Sc. (Hons) Degree in Information Technology Specialization in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

May 2023

Mobile Base Sinhala Book Reader for Visually Impaired Individuals

Project ID: 2023-198

Project Proposal Report

B.Sc. (Hons) Degree in Information Technology Specialization in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

May 2023

Declaration of The Candidate & Supervisor

I declare at this moment that the proposal I am presenting is entirely my work, and I have not incorporated, without proper acknowledgment, any material previously submitted for a degree or diploma at any other university or institute of higher learning. This proposal does not contain any material previously published or written by another person, except where the appropriate acknowledgment has been made in the text.

I am aware of the potential consequences of academic dishonesty and plagiarism, and I am committed to upholding the principles of honesty and intellectual integrity in all my academic endeavors.

Student IT Number	Student Name	Signature
IT20146238	Jayathunga T.M.	Thirosh.

The Supervisor should certify the proposal report with the following declaration.

The candidates mentioned above are currently conducting research for their undergraduate dissertation, under my supervision. As their supervisor, I certify this proposal report.

Signature of the Supervisor .... Date 04/05/2023

Abstract

Given the growth of technology, education has become increasingly important to all age groups. Currently, every sector of the world is developing with Artificial Intelligence (AI) at an unimaginable speed. With this, education is not limited to books, it is being developed with technology to make it easier to learn. Reading books can be the best way for a person to acquire the knowledge he needs. However, visually impaired people can read Braille and nowadays it is possible to do it through Computerized Braille systems or Text to Speech (TTS) systems. Although these types of technologies make things available in English easily accessible to visually impaired people, the equipment available in Sinhala is limited. Furthermore, such devices are very expensive in Sri Lanka, and it is not easy for a visually impaired person to buy these devices and they must spend a lot of money for them. Since Android is the most popular operating system in Sri Lanka, we decided that it would be better to make a mobile application based on it.

The accessibility of printed books and other publications in the Sinhala language for blind people in Lanka is not expanded. The main objective of this project is to enable them to have easy access to books and documents. Here, a visually impaired person navigates the app as needed by voice navigation, and the application gives the necessary commands to the visually impaired person. Here the mobile camera performs the necessary commands to accurately capture all the characters on a page of a Sinhala book. After that, optical character recognition separates and identifies the characters from the image with the corresponding characters. The Festival Synthesizing Framework's Text to Speech feature then allows the corresponding visually impaired individual to hear clearly. Here he can increase or decrease the speed of reading the book according to the commands. We intended to develop this Mobile Base Sinhala Book Reader to make it easier for visually impaired people by using audio recommendations and Sinhala voice notifications. We anticipate that this would boost the visually challenged community in Sri Lanka's delight in reading Sinhala novels.

Key Words: Text to Speech Sinhala, Natural Language Processing, Visually Impaired Individuals, Festival Synthesizing Framework, Speech Synthesis

Table of Content

List of Tables

List of Figures

List of Abbreviations

TTS – Text-to-Speech

OCR – Optical Character Recognition

NLP – Natural Language Processing

WHO - World Health Organization

CART – Classification and Regression Tree

DSP – Digital Signal Processing

1. Introduction

1.1. Background

There are many people who have vision problems, and all too frequently they go untreated. Globally, More than two billion people worldwide suffer from some form of vision problems, and of these, a minimum of one billion individuals have a condition that might have been avoided or is still unaddressed [1].

In the 21st century, knowledge has come to be seen as a necessity for successful existence. Reading becomes the primary way to learn new things. The difficulty for those who are blind or visually impaired is that there are many sources of knowledge that are not available in braille format, despite their eagerness to learn more. Text-to-speech and computer-assisted braille systems are just two of the technological remedies that have been created in response to this. We anticipate that this would boost the visually challenged community in Sri Lanka's delight in reading Sinhala novels.

Text-to-speech systems and computer-assisted braille systems are just two of the technological remedies that have been created in response to this. Most of these systems are expensive and broadly accessible only in industrialized nations [2]. Additionally, most text-to-speech tools are for the English tongue. However, these amenities are not available to Sri Lankans whose native language is Sinhala and whose level of English literacy is poor. As a result, we require such a mechanism.

Mobile-based technologies are quickly getting acceptance across the globe in the era of modern technological advancements. Even though Sri Lanka is still a developing country, most of its citizens own a smartphone [2]. The international trend also reveals that Android devices predominate the mobile phone market. Therefore, the most effective strategy to handle the problems encountered by visually disabled people in Sri Lanka would be a smartphone solution built on Android.

In conclusion, Android-based mobile book readers offer several benefits to Sri Lankans who are blind. Programmers can create specific apps with compatibility with the Sinhala language and accessibility features because of the open-source platform's versatility. Making mobile book readers for Android is a common option due to the platform's popularity and the accessibility of developer tools. Android-based mobile book readers can reach a larger population of people who are blind because Sri Lanka has a high percentage of Android users.

1.2. Literature Review

1.2.1. Text-to-Speech (TTS) Synthesizer

The study “NLP Applications of Sinhala: TTS and OCR” describes the creation of a Sinhala OCR which is the first published OCR scheme for the Sinhala language [3]. Visually impaired individuals in Sri Lanka face significant challenges in accessing written materials due to the lack of accessible formats. Braille and audiobooks are examples of traditional techniques that have been around for a while, but their accessibility has been restricted by factors like price, distribution, and availability. In recent years, text-to-speech mobile book readers have come to light as a viable answer to this issue, providing a practical and cost-effective replacement for conventional approaches. With an emphasis on Text-to-Speech (TTS) technology, the goal of this literature study is to determine the present level of knowledge on mobile-based Sinhala book readers for those who are visually impaired in Sri Lanka.

The goal of this review of the literature is to examine the level of knowledge of Text-to-Speech (TTS) technology for Sri Lankans who are visually impaired regarding mobile-based Sinhala book readers. Since most people in Sri Lanka speak Sinhala as their first language, it is essential to create mobile book readers that can read Sinhala text so that people who are blind can access written material [4]. By doing this, it is feasible to dramatically improve the quality of life for those who are blind or visually impaired and provide them independent access to written information.

The open-source Festival TTS system was used in the study on period modeling in TTS for Malayalam, and a machine learning approach known as CART or Classification and Regression Tree was used to construct data-driven phoneme length simulation. The significance of the phoneme in the sentence's context, its location, and the phonemes that come before and after it were among the factors used to estimate the duration of phonemes. The study researched the root mean square error (RMSE) between the real and predicted durations of time in an objective evaluation test to assess the comprehensibility of the synthesized speech [5]. After being objectively assessed, the model's performance was discovered to have an RMSE of 0.1188, showing a high degree of accuracy in predicting the phoneme length. The correlation between the actual and predicted phoneme lengths was also examined in the study, and it was found to be 0.9918, demonstrating a good link between the two. Overall, this research study showed how well data-driven methods for phoneme length modeling in Malayalam TTS systems worked. The study's conclusions may aid in boosting the naturalness and quality of synthetic Malayalam speech, which would eventually improve the user experience in programs like voice assistants, audiobooks, and other speech-based interfaces.

Many languages, including English and Sinhala, can have written text translated into speech using text-to-speech (TTS) systems. When compared to TTS systems created in other languages, English TTS systems are more widely available and perform better [6]. TTS system performance, however, varies depending on the language. This is because greater resources have been devoted to the development of English because it is a language that is extensively used.

The output of text-to-speech (TTS) systems can be audible text input. For people who are illiterate and visually challenged, they are especially crucial. TTS systems were initially created to serve the requirements of the community of people who were blind [7], but they now have a far wider variety of uses. For instance, they are employed as voice-based navigators, virtual assistants, and in public address systems. For these varied applications, a TTS system with good comprehension is essential.

The front end and the back end are the two primary parts of TTS. Text normalization and pre-processing are the front end's two primary responsibilities. The texts are marked into prosodic units such as phrases, sentences, and clauses after the text-to-phoneme conversion, which gives each word a phonetic transcription. A symbolic language representation made up of prosody information and phonetic transcriptions is the output of the front end. The synthesizer at the back end then transforms the output from the front end into sound [8]. As a high-level overview, the following graphic shows a typical TTS system.

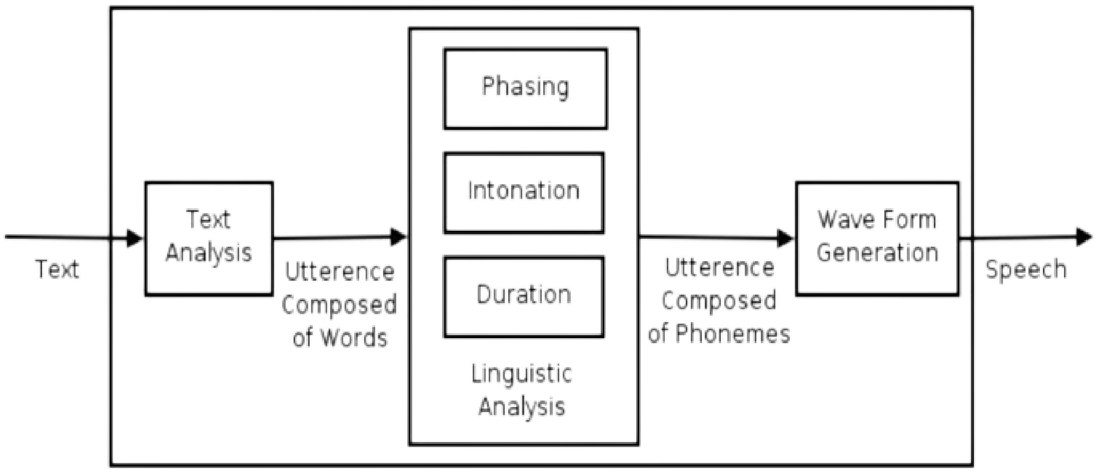

Fig. 1. Overview of the Typically TTS Diagram

Only a few research have been done thus far in Sri Lanka, where the creation of synthetic Sinhala voices is still in its infancy. The "Bhashitha" [9] project, an Android-based platform that offers reading solutions for Tamil, English, and Sinhala languages, has been created to fill this need. The Bhashitha project is a crucial milestone in Sri Lanka's effort to create synthetic voices for the languages of Sinhala, Tamil, and English.

The accessibility of textual materials for people with visual impairments in Sri Lanka is greatly influenced by the availability and functionality of TTS systems. Despite the recent development of multiple Sinhala TTS systems, they still perform worse than English TTS systems. We can determine the areas where Sinhala TTS systems require development and offer suggestions on how to enhance TTS systems for the Sinhala language by contrasting the performance of English and Sinhala TTS systems.

1.2.2. Android Base Related Work

The use of mobile devices has grown considerably in recent years, giving people a more convenient method to obtain information. For those who are blind, there is an increasing demand for portable text readers in Sri Lanka. The popularity of Android-based apps has increased because of the open-source status of the Android operating system and the availability of numerous developer tools [10]. This review of the literature aims to shed light on the advantages of creating Android-based mobile applications for visually impaired people in Sri Lanka that are based in Sinhala. The reason why we have created a mobile application related to the Android operating system can be seen in the diagram below.

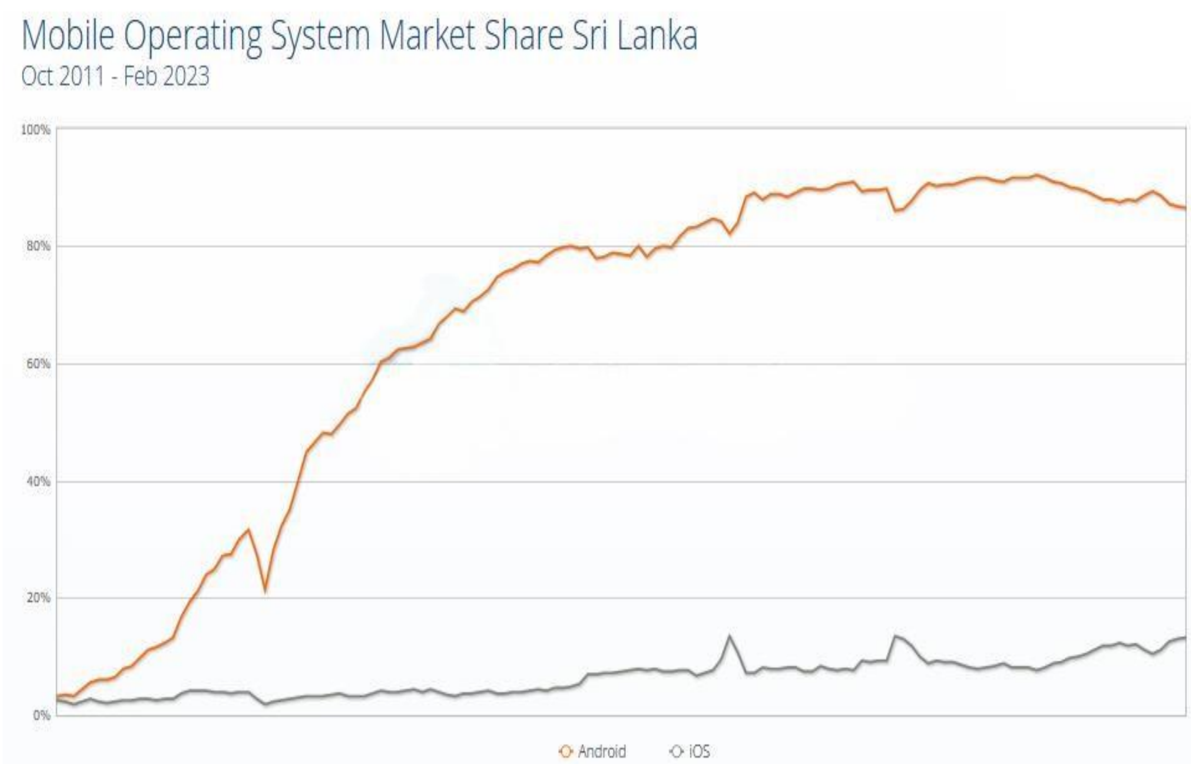

Fig. 2. Android and IOS OS Market in Sri Lanka

The goal of this project is to create a tool that would enable Sri Lankans who are blind to read books and other written materials in Sinhala. To illustrate the suggested method, a mobile application for the Android operating system is created. It can utilize the camera on a mobile device to capture a document, extract the text in it, and read it out to the user in Sinhala. The overview of the suggested remedy is depicted in the below image.

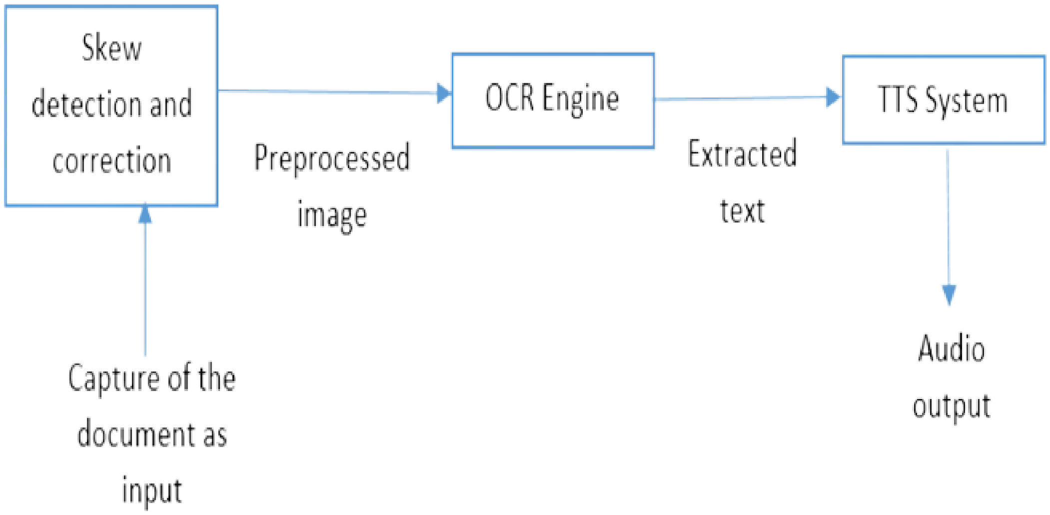

Fig. 3. Android Application Overview Diagram

Due to the open-source nature of the platform, which enables developers to build specialized applications suited to requirements, Android-based apps have grown in popularity among developers. Developers can build apps that are easier for people who are blind to use thanks to this flexibility [11]. To enhance the user experience, the application might incorporate disability features such as text-to-speech capabilities.

Android-based applications offer a high degree of customization, allowing developers to create applications that meet the specific needs of visually impaired individuals in Sri Lanka. For example, developers can create applications that support the Sinhala language and integrate text-to-speech features that improve the user experience.

The large developer community for Android offers tools so that programmers can make apps that are usable by people with vision impairments. Developers can use these materials to build mobile book devices that support the Sinhala language and accessibility features. Additionally, freely available frameworks and tools are part of these resources.

2. Research Gap

According to the literature review, there are several mobile applications for Text-to-Speech (TTS). Researchers have used a variety of methods, such as image processing and machine learning, to recognize writing and turn it into voice. To satisfy the requirements of Sri Lankans, there are currently no apps accessible that are specially designed with Sinhala Text-to-Speech (TTS) capabilities. For people and groups who depend on these services daily, this can be a significant restriction. But the available Text-to-Speech (TTS) mobile apps for the Sinhala language are limited and only a few are included in the table below.

TABLE 1 COMPARISON BETWEEN EXISTING SYSTEMS

Application Reference	Research A	Research B	Research C	Research D	Proposed System
High accurate Sinhala TTS conversion system	✗	✗	✓	✓	✓
OCR Text transferred to a TTS synthesizer in real time	✗	✓	✓	✗	✓
Adjust the Book reading Speed	✓	✓	✗	✗	✓
Support for Sinhala Language	✓	✗	✗	✓	✓
Android Mobile Application	✗	✗	✗	✗	✓

Natural language processing (NLP) is a component of a text-to-speech (TTS) generator known as the front end. The task of creating a symbolic depiction of the text to be considered depends on this NLP function. This depiction combines the words' phonetic transcription along with the intended tone and cadence, or prosody. The digital signal processing (DSP, often known for its backend) module comes after the natural language processing (NLP) [4] module and converts the visual data it takes into voice.

The creation of a very accurate Sinhala TTS conversion system is one of the most important research topics in mobile book readers for those with visual impairments. The national language of Sri Lanka is Sinhala, hence it's crucial to have a trustworthy TTS system that can faithfully translate written text into spoken Sinhala. This field of study focuses on creating sophisticated models and algorithms that can faithfully translate Sinhala text into voice with the proper emphasis, intonation, and pronunciation. Individuals who are blind or visually impaired may find it easier to access books and other written materials in the Sinhala language with the aid of a high-accuracy TTS conversion system.

The real-time transmission of OCR text to a TTS synthesizer is a key study field in mobile book readers for those with visual impairments. Utilizing OCR (Optical Character Recognition) technology [12], text can be extracted from images and converted into digital text that TTS systems can read. The goal of this research is to create an efficient and accurate OCR system that can extract text from photos of varying quality. The text must be sent in real-time to a TTS synthesizer after the OCR system has identified it to create spoken words. For those who are blind or visually challenged, this study field is essential since it enables them to access written material and instantly translate it into voice.

The ability to change the reading pace of books is one of the key characteristics of mobile book readers for those who are blind or visually impaired. The goal of this research is to provide a user-friendly interface that enables users to choose their preferred place for the TTS system. Because different users may have different reading needs, this feature enables them to customize the reading speed to suit those needs. In this field of study, algorithms are designed and created that allow the TTS system's reading speed to be changed without compromising the accuracy of the speech.

The creation of an Android mobile application and support for the Sinhala language is required for the creation of mobile book readers for people with vision impairments. Creating an application that is simple to download and install on Android smartphones is the focus of this study topic. Visually challenged users must be able to easily use the application's user-friendly interface. The program also must support the Sinhala language and include TTS, OCR, and book-reading features. This research involves creating a strong application that can give blind people simple access to books and other written material in the Sinhala language.

3. Research Problem

According to the World Health Organization's (WHO), there are currently at least 290 million visually impaired persons worldwide, including almost forty million blind people [1]. For those who are visually impaired individuals, reading books can be very difficult because they commonly cannot obtain printed materials in a way that is convenient for them. Many books are still not readily available in a way that the blind can read despite advancements in assistive technology such text-to-speech software and Braille displays. For blind people who want to increase their knowledge, participate in literary encounters, and enhance their educational and job possibilities, this restricted accessibility can present significant hurdles.

Another problem is the price of specialist equipment and software, which for many blind individuals can be unaffordable. Even if these tools are accessible, perusing a conventional paper book might not be a similar experience with them. For instance, braille displays can be clumsy and sluggish, and text-to-speech applications can have trouble with complicated English and formatting.

The lack of audiobooks and Braille resources is another problem [13]. Even though there are more audiobooks available, there are still few of them compared to the enormous number of written novels. Due to the lengthy and expensive procedure involved in converting written books into Braille, finding Braille books is even more difficult. This implies that the most recent bestsellers or widely used instructional resources may not be available to blind people.

In conclusion, a shortage of access to written materials is the primary obstacle that blind people encounter when perusing books. Even with advancements in assistive technology, there are still many obstacles to be overcome, including the price of specialized hardware and software, the scarcity of audiobooks and Braille materials, and the challenge of providing a reading experience that is comparable to that of conventional printed books. To overcome these obstacles, there must be a concerted effort to increase the accessibility of books for blind readers and to guarantee that they have the same access to reading and educational possibilities as sighted people.

4. Objectives

4.1. Main Objective

A Sinhala book reader for the visually impaired is a software program designed to make reading accessible for individuals with visual impairments. It combines various technologies to provide a seamless reading experience.

Most people in Sri Lanka use Sinhala as their native language, making it crucial to create mobile book readers that can read texts in Sinhala for those who are blind or visually impaired. By allowing visually impaired people to access textual information with ease and independence, this technology has the potential to greatly improve their quality of life. Due to its ability to transform the written text into voice, mobile book readers that use TTS have grown in popularity. This enables those who are blind or visually impaired to listen to information.

The two main parts of the program we intend to develop are a Text-to-Speech (TTS) synthesizer and an Optical Character Recognition (OCR) technology to identify the most common Sinhala characters [6]. The device utilizes optical character recognition (OCR) technology to convert the text from a physical book into a digital format. Then, a text-to-speech synthesizer reads the text out loud in the Sinhala language, making it easier for visually impaired users to follow along.

In addition to the text-to-speech synthesizer, the device also includes audible guidance to help navigate the app and identify the distance to the book being read. This makes it easier for visually impaired users to find their place in the book and keep track of their progress.

4.2. Sub Objective

4.2.1. Text-to-Speech (TTS)

The process of creating human-like speech mechanically is known as speech synthesis. A voice synthesizer is a machine that uses computers to make sounds that mimic real speech. Text to Speech transforms language text into speech, while other methods transform ways to represent symbolic linguistics [2].

Concatenation, which entails combining pre-recorded speech fragments kept in a database, is a popular technique for producing artificial speech around speech synthesis. Another strategy includes turning written text into speech with the Text-to-Speech (TTS) method, and some systems use various strategies to change symbolic language representations into voice. Speech synthesis uses two main technologies, formant, and concatenation, to create synthetic speech that closely resembles natural speech.

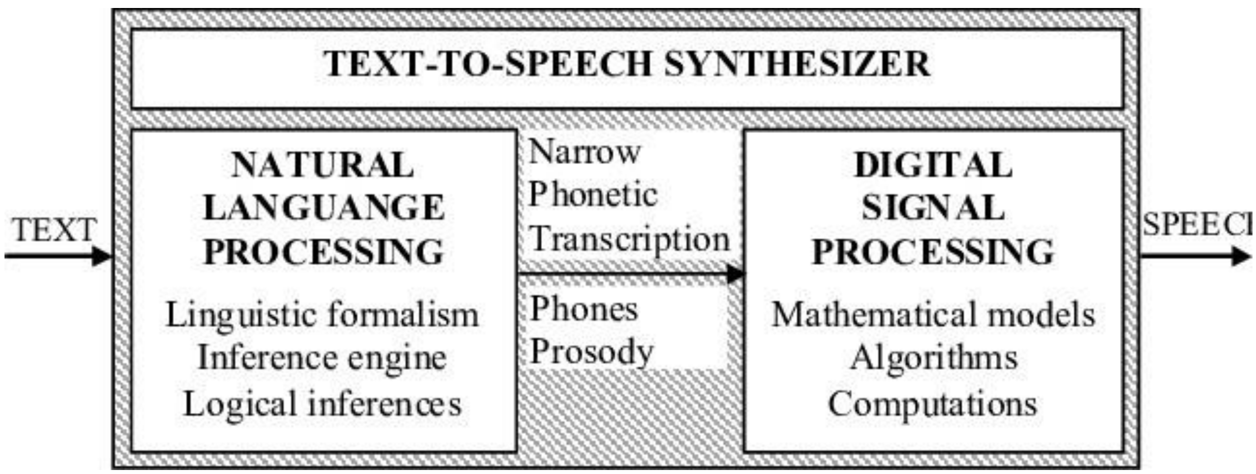

Fig. 4. General Functional Diagram for Text-to-Speech (TTS) Synthesizer

Text-to-speech (TTS) technology comprises two key components: the front end and the back end [14]. Text normalization and pre-processing are the two main duties of the front end. While the pre-processing stage recognizes the individual words in the text and transforms them into their corresponding phonetic transcriptions, the text normalization phase makes sure that the input text is in a standardized format.

The text is separated into prosodic units, such as phrases, sentences, and clauses, and is then transformed into phonemes. These prosodic units are tagged with pertinent prosody information. A symbolic language representation of the input text is subsequently created by the front-end using prosody information and phonetic transcriptions.

The symbolic language representation created by the front end is used at the back end to generate the synthesized sound [15]. By transforming the symbolic representation of the text into audible sound, a technique known as synthesis is used to do this.

4.2.2. Phonemes in Sinhala Language

➢ **Sinhala Consonant**

The 26 consonants of Sinhala, Sri Lanka's national language, give it a unique sound. Table of Figure 5 includes a list of these consonants, which constitute an important component of the language's phonetic inventory [16]. The retroflex consonants in Sinhala, which are spoken by curving the tongue back, and the hissing alveolar fricatives, are two distinctive characteristics that set them apart from other consonants. To further emphasize its distinctiveness, the language has a series of pre-nasalized voiced stops that are unique to Sinhala.

		Labial	Dental	Alveolar	Retroflex	Palatal	Velar	glottal
Stops	**Voiceless**	p	t		ʈ		k	
	Voiced	b	d		ɖ		g	
Affricates	**Voiceless**					c		
	Voiced					j		
Pre-nasalized voiced stops		b̃	d̃		ɖ̃		g̃	
Nasals		m		n		ɲ	ŋ	
Trill				r				
Lateral				l				
Spirants		f	s			s̀		h
Semivowels		v				y		

Fig. 5. Spoken Sinhala Consonant Classification

The sound patterns of Sinhala's consonants are fundamental to the language's identity and cultural legacy. As seen by the expanding usage of English loanwords, they constitute a crucial component in the language's development and have adjusted to contemporary culture.

➢ **Sinhala Vowels**

The 14 vowels in Sinhala are renowned for contributing to the distinctive sound of the language. These vowels, which are mentioned in Table of Figure 6, are a crucial component of the phonetic inventory of the language. The vowels of Sinhala contain distinctive characteristics that set them apart from those of other languages, such as a significant degree of phonemic length contrast and a sophisticated system of vowel harmony.

	Front		**Central**		**Back**	
	Short	long	Short	long	short	long
High	i	i:			u	u:
Mid	e	e:	ə	ə:	o	o:
Low	æ	æ:	a	a:		

Fig. 6. Spoken Sinhala Vowel Classification

The vowels and sounds of the Sinhala language reflect the language's rich cultural past. They continue to develop and adapt to contemporary circumstances while playing a significant role in the identity and legacy of the language. An illustration of how the language is integrating new sounds and words while keeping its own qualities is the rising usage of English loanwords.

Consonants and vowels together make up most of the language's sounds and identities in Sinhala. They continue to play a crucial part in the language's development and add to its rich cultural history.

4.2.3. Sinhala Character Set

The official language of Sri Lanka is Sinhala, which is comprised mostly of the Sinhala character set. Each of the 40 consonants and 20 vowels in the set has a distinctive characteristic that adds to the distinctive sound of the language [16]. By adding or deleting diacritics, which are symbols that change the sounds made by the consonants and vowels, it is possible to determine the composition of consonants and vowels.

There are a total of 18 diacritics in the Sinhala letter set, 17 of which are known as vowel modifiers that indicate vowels. These vowel modifiers are crucial for differentiating between the various vowels in the language, including nasalized vowels as well as long and short vowels. The last diacritical mark indicates the unaltered consonant form.

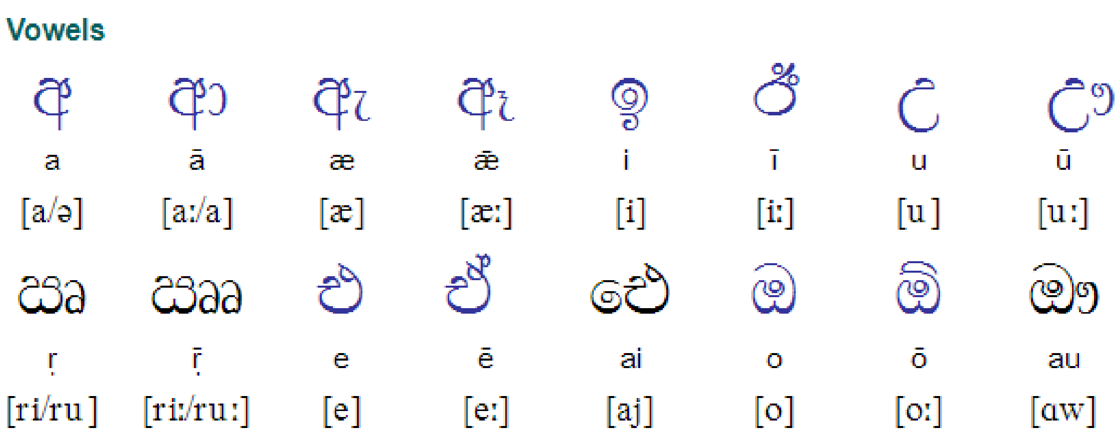

Fig. 7. Sinhala Language 20 Vowels Table

Vowels are a crucial part of the Sinhala language and contribute significantly to the distinctive sound of the language. The 20 vowels of the Sinhala letter set each have distinctive characteristics that add to the depth and complexity of the language. There are three types of vowels in the Sinhala language: short, long, and nasalized.

Simple vowel symbols are used in Sinhala to express short vowels, whereas diacritical pure is used to indicate long vowels. On the other hand, nasalized vowels are denoted by the addition of the niggahita tilde diacritic. To distinguish between various vowels in the Sinhala language, certain diacritics must be used.

The 40 consonants that make up the Sinhala letter set each have distinctive characteristics and sounds, making them an equally important part of the Sinhala language. Pure consonants and compound consonants are the two types of consonants used in the Sinhala language.

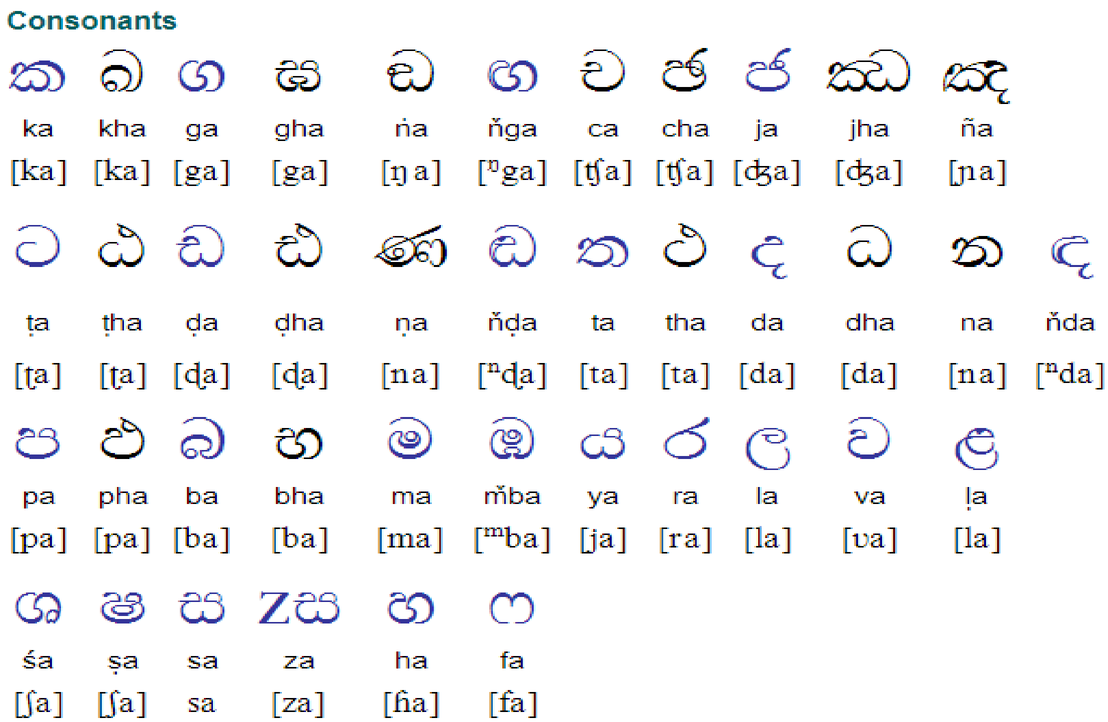

Fig. 8. Sinhala Language Consonants Table

Pure consonants, which are consonants devoid of vowels, stand in for straightforward linguistic sounds. Contrarily, compound consonants, which are made up of a mix of consonants and vowels, stand in for the language's more complex sounds. By adding or deleting diacritics from the consonants, it is possible to determine how the consonants and vowels are composed.

Finally, vowels and consonants both contribute significantly to the distinctive sound of the Sinhala language. The 40 consonants of the Sinhala letter set to reflect a broad variety of basic and complicated sounds, whilst the 20 vowels are crucial for differentiating between the many sounds in the language. Diacritics are crucial for showing consonant and vowel composition and for enhancing the depth and complexity of the Sinhala language. Overall, the accessibility and inclusiveness of studying and communicating in the Sinhala language for those who are visually impaired have been substantially enhanced by the availability of text-to-speech choices.

5. Methodology

5.1. Overall System Diagram

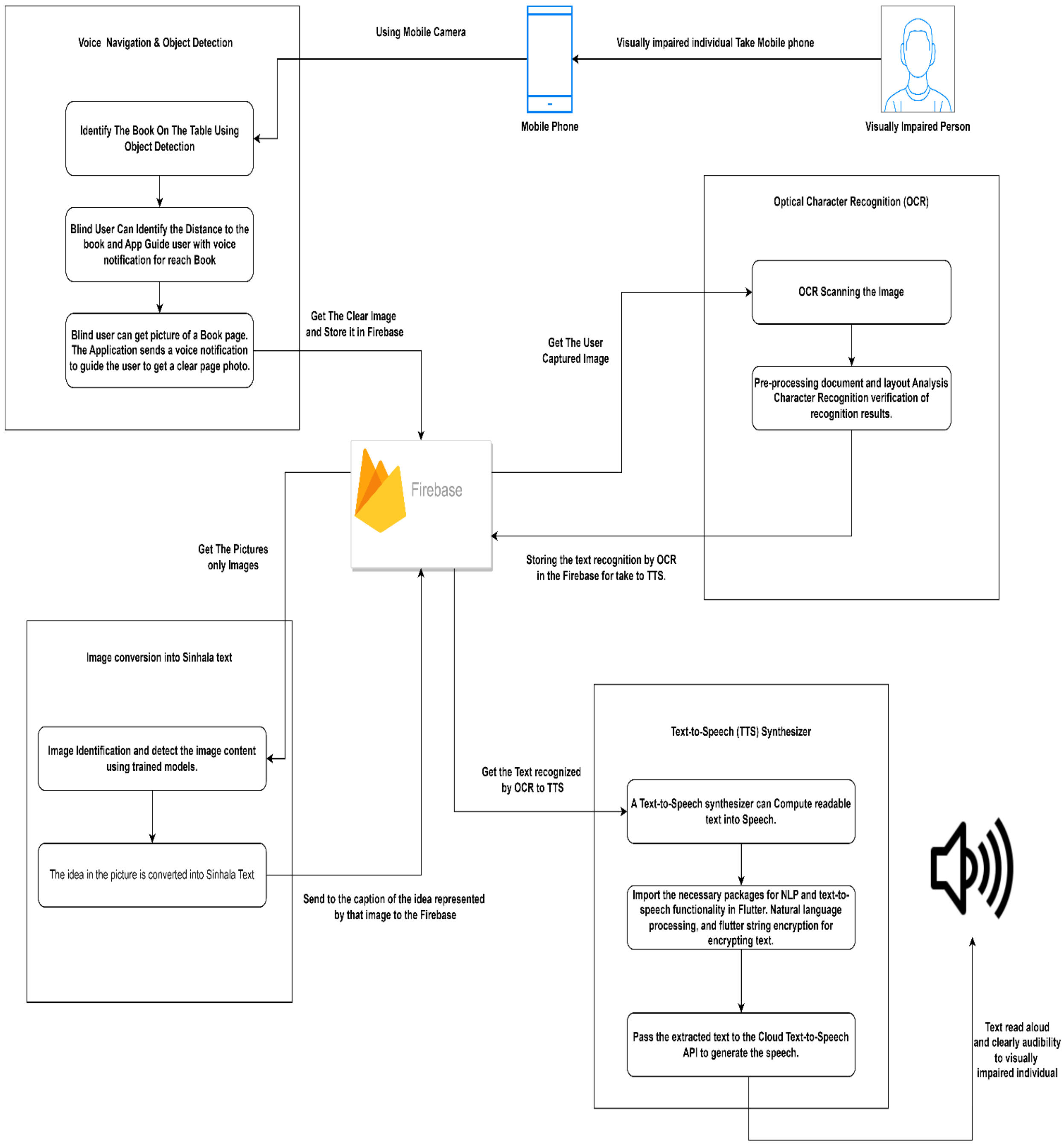

Fig. 9. Project Overall System Diagram

Audible guidance helps the user navigate the app through the functions of the app and get a clear idea and guidance whenever the user faces difficulty performing a task. When a user wants to find a book on the table, he opens the camera through the app and points it toward the table or desk and the app identified the user's hand and navigates to the book. Using real-time Image processing technology identifying the dangerous object near the blind user and the distance to the object will notify verbally to the user and identify the probability of occurring an accident the user. Users will navigate in a pristine environment evading dangers and harmful things.

Sinhala Character Identification and word formation through the engine and translate to the speech then send to the TTS [17]. Users can use a background process to read the time even when the app is not open. When the user launches the software using voice instructions, the program ought to have been ready to start up the camera as well. The program should be able to quickly scan the document when the user launches it. The program must be able to detect and focus on the paper in front of the camera. Until the document is within the capture frame, the system should alert the user audibly. The software notifies the user and guides them through the process of taking a picture of a page in a book using their phone's camera. The device's storage should be used to store the image that was captured. Before submitting information to the OCR system [18], the system must detect and correct skew.

A computer vision system that uses cutting-edge methods to recognize and describe objects and scenes in real-time is an image detection software for blind students. The program takes pictures of the user's surroundings using the camera on a smartphone or tablet [19]. These images are subsequently processed using image-processing techniques including edge detection, color analysis, and feature extraction to identify the objects and their attributes. Machine learning models are used to identify and categorize the objects in the image once the objects can be in the image using object detection methods [11]. These models use methods like convolutional neural networks (CNNs) to recognize objects in real-time and are trained on massive datasets of annotated photos.

Here, the main purpose of using Text to speech (TTS) technology is to give a blind person the ability to access the written text of a Sinhala book [14]. This allows them to easily listen to the valuable content of Sinhala books. TTS technology allows the written text in a Sinhala book to be read out loud in a natural-sounding voice, which makes it easier for visually impaired people to understand the content. The technology uses computer algorithms [20] to analyze the Sinhala text and generate an appropriate pronunciation, intonation, and rhythm for each word and sentence.

5.2. Individual System Diagram

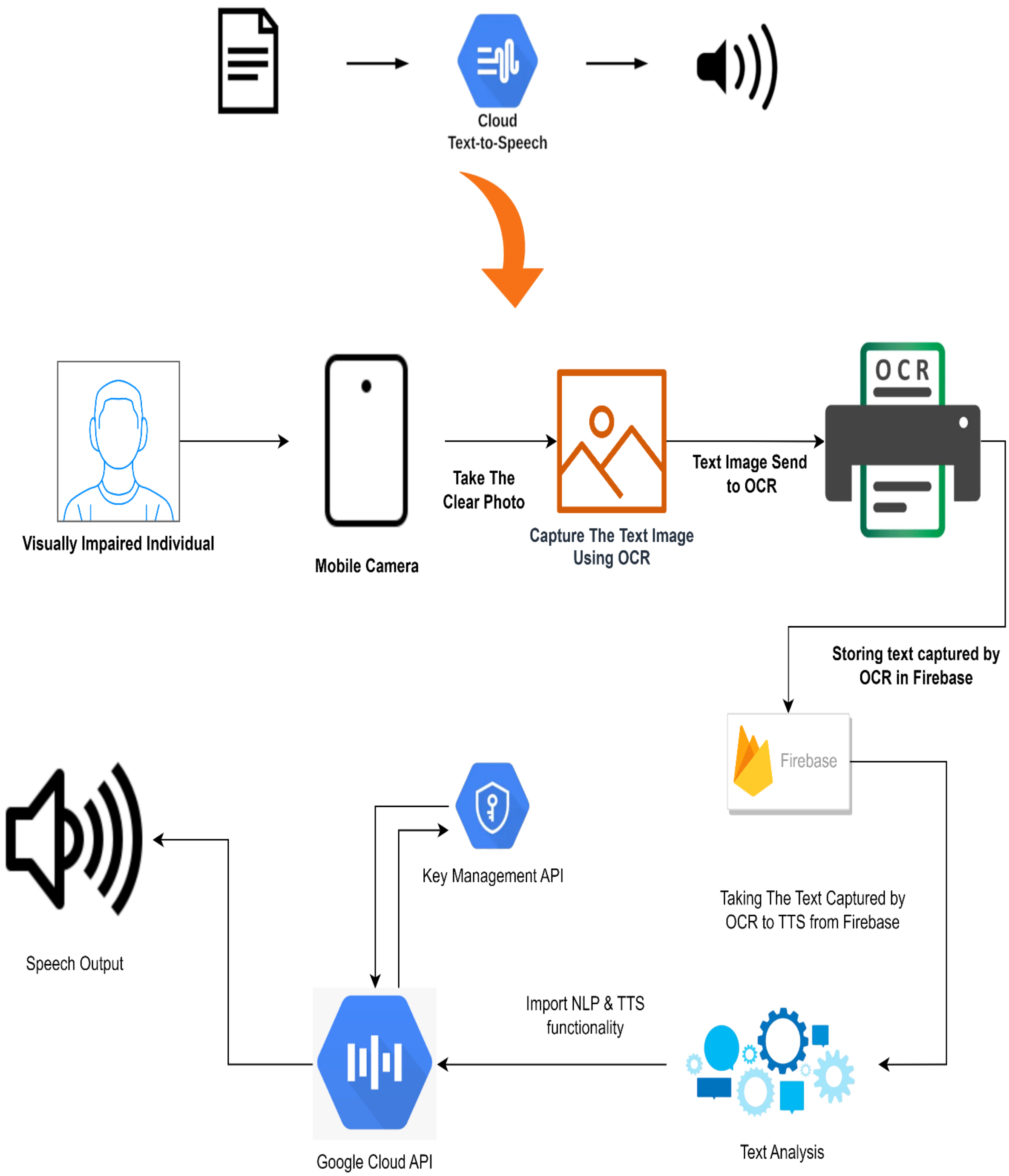

Fig. 10. Text-to-Speech (TTS)

A computer-based system called a text-to-speech synthesizer can turn text that can be read by computers into speech. Many crucial processes are involved in the translation of text into speech. Text analysis, linguistic analysis, and wave-form production are the three basic stages that can be distinguished between these processes [12].

The procedures are as follows:

- When OCR is finished, a visually impaired person should receive an audio notification.
- At this point, transferred to the TTS system should be texts that have been extracted and identified in Sinhala characters.
- Lastly, TTS should be utilized to read the Sinhala text identified in the camera image out and clearly audible to persons who are blind.
- Overall, the primary goal of using TTS in the Sinhala language book reader for people who are blind is to give them a way to access written material and appreciate the content of books despite their visual impairment.

5.3. Commercialization of the Project

- Identifying the Target Audience: In this case, the target audience would be visually Impaired Individuals who could benefit from the features of the Mobile Application.
- Revenue Generation: Commercialization also involves generating revenue from the mobile application. This could involve charging for the application itself, offering premium features for a fee.
- Promotions: The target audience should be reached through targeted campaigns on social media and other channels. Collaboration with organizations that work with visually impaired individuals, such as libraries or schools, can also be an effective way to reach the target audience.

6. Software Specifications

6.1. Functional Requirements

- The mobile book reader must have a high-quality text-to-speech (TTS) synthesizer capable of accurately reading Sinhala text out loud.
- Individuals who are blind or visually impaired must be able to customize the TTS function's reading speed settings to suit their preferences.
- The TTS function should have a pause and resume feature so that the user can stop and start the reading at any time.

6.2. Non-Functional Requirements

- Accessibility: The program needs to be created with people who are visually impaired individuals in mind.
- Usability: The program must be simple to use, have simple controls, and provide users with straightforward directions.
- TTS voice quality: The application's Text-to-Speech (TTS) synthesizer must generate high-quality, believable speech.
- Performance: Even on mid-range devices, the program must be prompt and quick and Responsive and seamless TTS conversion and navigation.
- Security: The application should be secure to protect user information and reading history.
- Reliability: The application should be reliable and provide consistent TTS conversion.

6.3. Tools and Technology

- Android Studio
- Flutter
- Firebase
- TTS Engine
- Python Flask
- Amazon Polly
- Google API
- TensorFlow
- NLP

7. Work Breakdown Structure

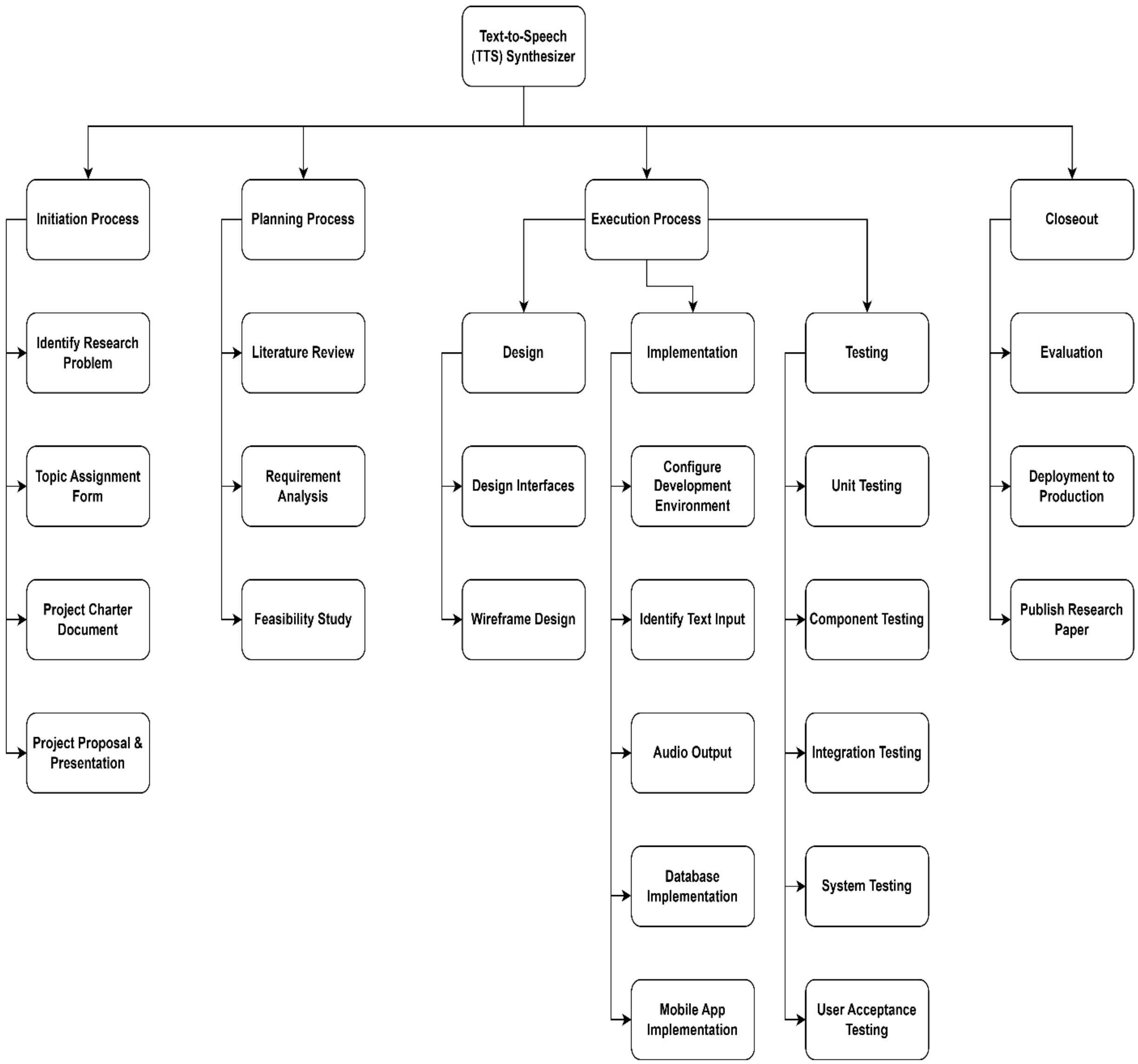

Fig. 11. Work Breakdown Chart

8. Gantt Chart

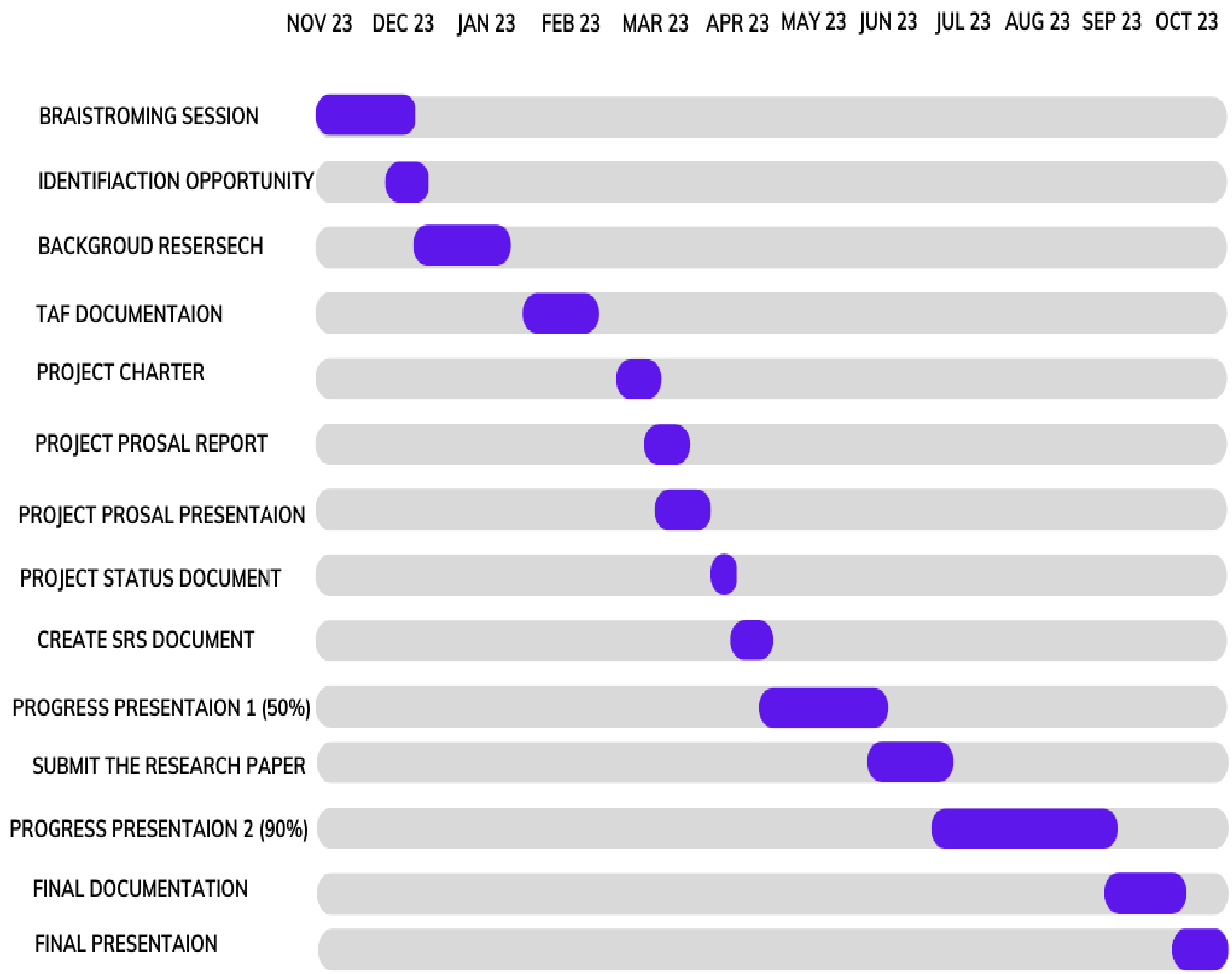

Fig. 12. Gantt Chart

9. References

[1] WHO, *World report on vision*, vol. 214, no. 14. 2019. [Online]. Available: https://www.who.int/publications-detail/world-report-on-vision

[2] A. Mishangi, “Android based sinhala document reader for visually impaired people,” 2021.

[3] R. No, “University of Kelaniya,” vol. 94, no. March, p. 2976600, 2016, [Online]. Available: http://www.kln.ac.lk/

[4] R. Carlson and B. Granström, “Speech Synthesis,” *Handb. Phonetic Sci. Second Ed.*, pp. 781–803, 2010, doi: 10.1002/9781444317251.ch21.

[5] B. P. Tóth, P. Nagy, and G. Németh, “Towards Modeling Interrogative Sentences in HMM-based Speech Synthesis,” *The Phonetician*, vol. 109–101, no. 2014-I–II, pp. 24–42, 2014.

[6] D. H. Klatt, “Review of text-to-speech conversion for English,” *J. Acoust. Soc. Am.*, vol. 82, no. 3, pp. 737–793, 1987, doi: 10.1121/1.395275.

[7] A. Joy* and D. R. Saranya, “A Pilot Research on Android Based Voice Recognition Application,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 4, pp. 7272–7277, 2019, doi: 10.35940/ijrte.d5284.118419.

[8] K. Wasala, A., Weerasinghe, R., Gamage, “A Sinhala Text-to-Speech System,” 2007.

[9] D. S. S. De Zoysa, J. M. Sampath, E. M. P. De Seram, D. M. I. D. Dissanayake, L. Wijerathna, and S. Thelijjagoda, “Project Bhashitha - Mobile based optical character recognition and text-to-speech system,” *13th Int. Conf. Comput. Sci. Educ. ICCSE 2018*, no. Iccse, pp. 623–628, 2018, doi: 10.1109/ICCSE.2018.8468858.

[10] M. Awad, J. El Haddad, E. Khneisser, T. Mahmoud, E. Yaacoub, and M. Malli, “Intelligent eye: A mobile application for assisting blind people,” *2018 IEEE Middle East North Africa Commun. Conf. MENACOMM 2018*, no. September, pp. 1–6, 2018, doi: 10.1109/MENACOMM.2018.8371005.

[11] P. Kardyś, A. Dabrowski, M. Iwanowski, and D. Huderek, “A new Android application for blind and visually impaired people,” *Signal Process. - Algorithms, Archit. Arrange. Appl. Conf. Proceedings, SPA*, pp. 152–155, 2016, doi: 10.1109/SPA.2016.7763604.

[12] R. Weerasinghe, A. Wasala, D. Herath, and V. Welgama, “NLP applications of Sinhala: TTS & OCR,” *IJCNLP 2008 - 3rd Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, vol. 2, pp. 963–966, 2008.

[13] G. V. S. Murthy *et al.*, “The Sri Lanka National Blindness, Visual Impairment and Disability Survey: rationale, objectives and detailed methodology,” *Ceylon Med. J.*, vol. 63, no. 5, p. 3, 2018, doi: 10.4038/cmj.v63i5.8735.

[14] K. Sodimana *et al.*, “A Step-by-Step Process for Building TTS Voices Using Open Source Data and Frameworks for Bangla, Javanese, Khmer, Nepali, Sinhala, and Sundanese,” *6th*

Work. Spok. Lang. Technol. Under-Resourced Lang. SLTU 2018, no. August, pp. 66–70, 2018, doi: 10.21437/SLTU.2018-14.

[15] A. A. Kumar, B. Senthilvasudevan, and H. U. Farhan, "Translation of Multilingual Text into Speech for Visually Impaired Person," *7th Int. Conf. Commun. Electron. Syst. ICCES 2022 - Proc.*, no. Icces, pp. 60–64, 2022, doi: 10.1109/ICCES54183.2022.9835819.

[16] A. Wasala, R. Weerasinghe, and K. Gamage, "Sinhala grapheme-to-phoneme conversion and rules for Schwa epenthesis," *COLING/ACL 2006 - 21st Int. Conf. Comput. Linguist. 44th Annu. Meet. Assoc. Comput. Linguist. Proc. Main Conf. Poster Sess.*, no. July, pp. 890–897, 2006, doi: 10.3115/1273073.1273187.

[17] L. Nanayakkara, C. Liyanage, P. T. Viswakula, T. Nadungodage, R. Pushpananda, and R. Weerasinghe, "A Human Quality Text to Speech System for Sinhala," *6th Work. Spok. Lang. Technol. Under-Resourced Lang. SLTU 2018*, no. August, pp. 157–161, 2018, doi: 10.21437/SLTU.2018-33.

[18] N. A. Azis, T. V. Tjahja, and A. S. Nugroho, "Evaluation of Text-to-Speech Synthesizer Unpredictable Sentences Test : IndoTTS , eSpeak , and," no. April 2014, 2011.

[19] A. Walczak and A. Szarkowska, "Text-to-speech audio description for visually impaired children," no. March, pp. 24–25, 2011.

[20] P. Jayawardhana, A. Aponso, N. Krishnarajah, and A. Rathnayake, "An Intelligent Approach of Text-To-Speech Synthesizers for English and Sinhala Languages," *2019 IEEE 2nd Int. Conf. Inf. Comput. Technol. ICICT 2019*, no. May, pp. 229–234, 2019, doi: 10.1109/INFOCT.2019.8711051.

Appendices

Figma UI: https://www.figma.com/file/vejcciCAR9DNNTkbZuBoWr/AudioSight?node-id=0%3A1&t=q57j4R1OdAYZhRPT-1

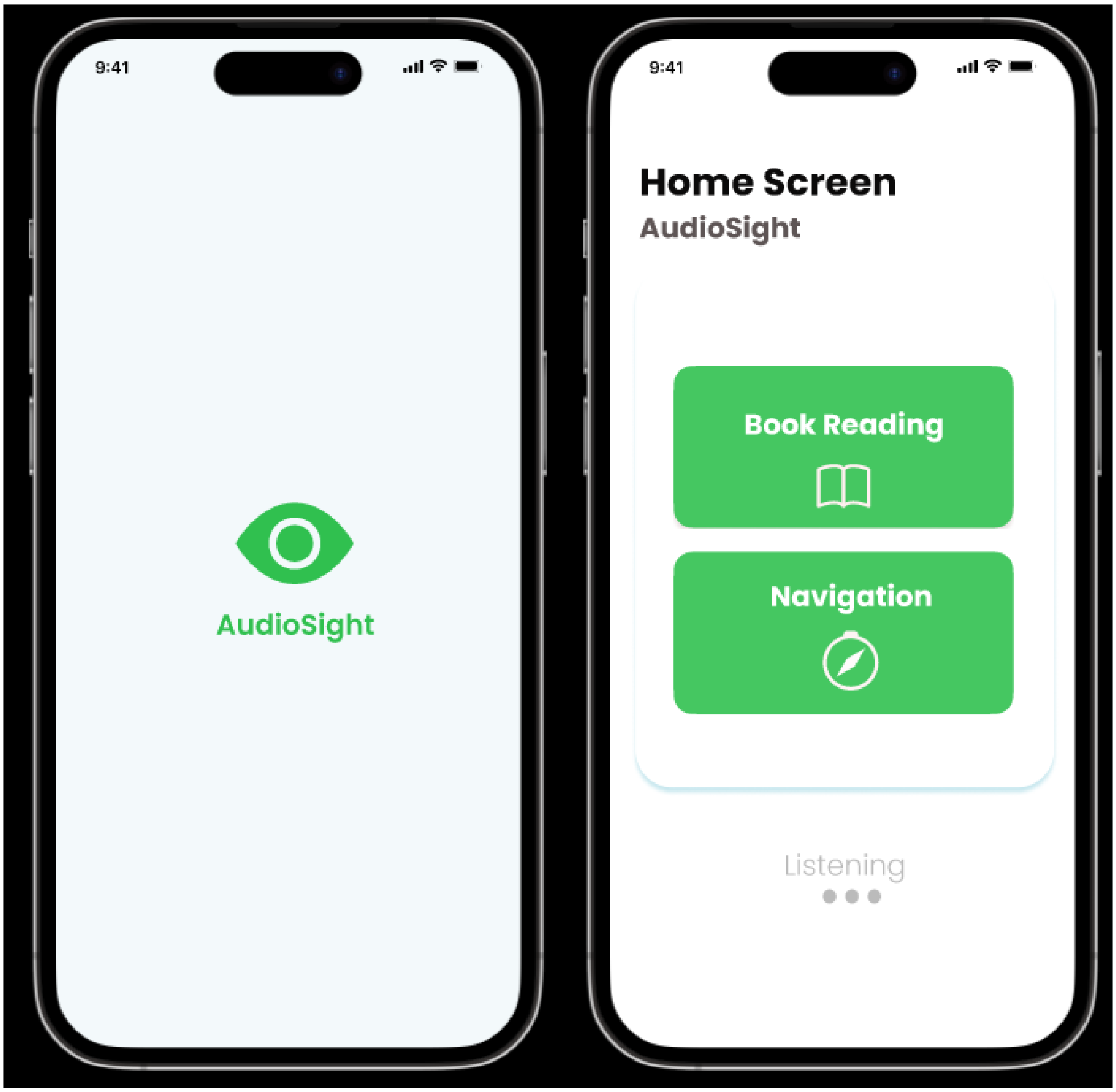

Fig. 13. Figma UI Designs


9:41
Navigation
Instruction
Reading