Enhancing the Communication Experience for the Deaf and Hard-of-Hearing using AI Language Models

By

Githmi Niseka De Silva - Registration No: 2020/IS/026

N.G.A.N Bandara - Registration No: 2020/IS/015

M.S.F. Shimra - Registration No: 2020/IS/101

Supervisor : Dr. K.A.K.T Karunanayaka

Co-supervisor : Mrs. S.S Thrimahavithana

This dissertation is submitted to the University of Colombo School of Computing
In partial fulfillment of the requirements for the
Degree of Bachelor of Science Honours in Information Systems

University of Colombo School of Computing
35, Reid Avenue, Colombo 07,
Sri Lanka

April 2025

Declaration

We, Githmi Niseka De Silva (2020/IS/026), N.G.A.N Bandara (2020/IS/015), and M.S.F. Shimra (2020/IS/101) hereby certify that this dissertation entitled Enhancing the Communication Experience for the Deaf and Hard-of-Hearing using AI Language Models is entirely our own work and it has never been submitted nor is currently being submitted for any other degree.

Candidate Name: Githmi Niseka De Silva

....

Signature of Candidate Date : 2025.06.30

Candidate Name: N.G.A.N Bandara

....

Signature of Candidate Date : 2025.06.30

Candidate Name: M.S.F. Shimra

....

Signature of Candidate Date : 2025.06.30

I, Dr. K.A.K.T Karunanayaka, certify that I supervised this dissertation entitled Enhancing the Communication Experience for the Deaf and Hard-of-Hearing using AI Language Models conducted by Githmi Niseka De Silva, N.G.A.N Bandara, and M.S.F. Shimra in partial fulfillment of the requirements for the degree of Bachelor of Science Honours in Information Systems.

....

Signature of Supervisor Date : 2025.06.30

I, Mrs. S.S Thrimahavithana, certify that I supervised this dissertation entitled Enhancing the Communication Experience for the Deaf and Hard-of-Hearing using AI Language Model conducted by Githmi Niseka De Silva, N.G.A.N Bandara, and M.S.F. Shimra in partial fulfillment of the requirements for the degree of Bachelor of Science Honours in Information Systems.

. . . .

Signature of Co-Supervisor Date :

Abstract

Effective communication in healthcare is critical for accurate diagnosis and treatment, but deaf and hard-of-hearing (DHH) people experience substantial challenges in medical consultations. This problem grows more significantly in Sri Lanka, since there are few assistive tools accessible for Sinhala-speaking DHH patients. The identified research gap is the absence of accessible, real-time communication tools that enable seamless interaction between healthcare professionals and DHH patients in Sinhala. This research bridges the gap by developing a mobile application that enhances bidirectional communication in medical settings, addressing the user group's preference for mobile solutions and their need for support in healthcare contexts. The study focuses on creating and evaluating a smartphone application that allows healthcare professionals to speak in Sinhala, which is then transcribed into text. The app provides three contextually relevant responses for the DHH patient, who can choose, modify, and confirm one, which is then turned back into voice for the doctor. Furthermore, the system accepts text-based input from the patient, allowing doctors to answer verbally. The evaluation uses a mixed-method approach, integrating quantitative indicators like mobile application transcription accuracy and answer generating speed with qualitative feedback from interviews with DHH individuals. Usability testing was conducted with Deaf/Hard-of-Hearing (DHH) people and communication facilitators to evaluate accessibility, efficiency, and user satisfaction. The findings show that the application enhances communication clarity and lowers misconceptions during medical consultations. DHH users provided positive feedback, emphasizing the application's real-time response creation and ease of use. This research contributes to assistive technology for the DHH community in Sri Lanka by providing a realistic answer to a crucial healthcare accessibility issue.

Acknowledgement

We would like to express our sincere gratitude to everyone whose valuable initiatives were vital in completing this thesis.

First and foremost, we want to express our gratitude to our supervisor, Dr. K.A.K.T Karunanayaka, whose guidance, encouragement, and insightful perspectives have been essential to this research project. We sincerely appreciate the time and effort he invested. His expertise profoundly shaped the direction of this work, and the constant encouragement kept us motivated during challenging phases.

We also extend our sincere thanks to Mrs. S.S Thrimahavithana, our co-supervisor, for her guidance and support throughout our research.

We are immensely grateful to the Dr. Reijntjes School for the Deaf, "SAWAN SAHANA" DHH community members who participated in this study. Their feedback was instrumental in designing an application that truly addresses real-world communication barriers. Special thanks to Mr. Muditha for his time, expertise, and willingness to test the application in clinical settings.

We also want to express our appreciation to the writers whose writings have provided the thesis with essential references. Their academic works have offered crucial viewpoints and ideas that have improved our comprehension of the topic.

Finally, we would like to thank our family and friends for their constant encouragement and support during this journey.

We sincerely thank everyone listed and the countless others who have helped us along the way for their invaluable efforts.

Table of Contents

List of Figures

List of Tables

List of Abbreviation

AI Artificial Intelligence

ALD Assistive Listening Devices

ASR Automatic Speech Recognition

DHH Deaf and Hard-of-Hearing

DSR Design Science Research

HCI Human-Computer Interaction

IS Information systems

LLM Large Language Model

NLP Natural Language Processing

SSL Sri Lankan Sign Language

STT Speech-To-Text

TTS Text-To-Speech

UCD User-Centered Desig

UI User Interface

WHO World Health Organization

Chapter 1

Introduction

Communication plays a central role in human interaction. However, for individuals who are deaf or hard of hearing (DHH), which refers to those with partial or complete hearing loss, this fundamental process can become a major challenge. While many in the DHH community use sign language, lip-reading, or written communication to interact with others, these methods often fall short in situations where quick and accurate communication is essential. Misunderstandings, lack of access to critical information, and feelings of social isolation are common outcomes [1].

This issue becomes even more serious in healthcare settings. Deaf and hard-of-hearing individuals frequently encounter difficulties when communicating with medical professionals who may not have the training or tools needed to interact effectively with them. As a result, important information about symptoms, treatments, or follow-up care can be misunderstood or missed entirely, which can negatively affect patient health outcomes [2]. In addition to these communication challenges, broader systemic issues such as a lack of awareness about the unique needs of DHH patients and limited access to real-time assistive tools further complicate their healthcare experiences [3].

Our research focuses on addressing this gap through a mobile application designed for deaf individuals to use in medical contexts, especially during one-on-one conversations with doctors. Traditionally, when a deaf individual visits a doctor, they might rely on a sign language interpreter, a family member, or lip-reading. These methods are not always reliable or private. Our app aims to provide a more independent, accurate, and real-time communication experience without the need for external assistance.

The communication process within the application is designed to ensure seamless, two-way interaction between Deaf/Hard-of-Hearing (DHH) users and healthcare providers. The following flowchart illustrates the operational flow of the application during a typical conversation:

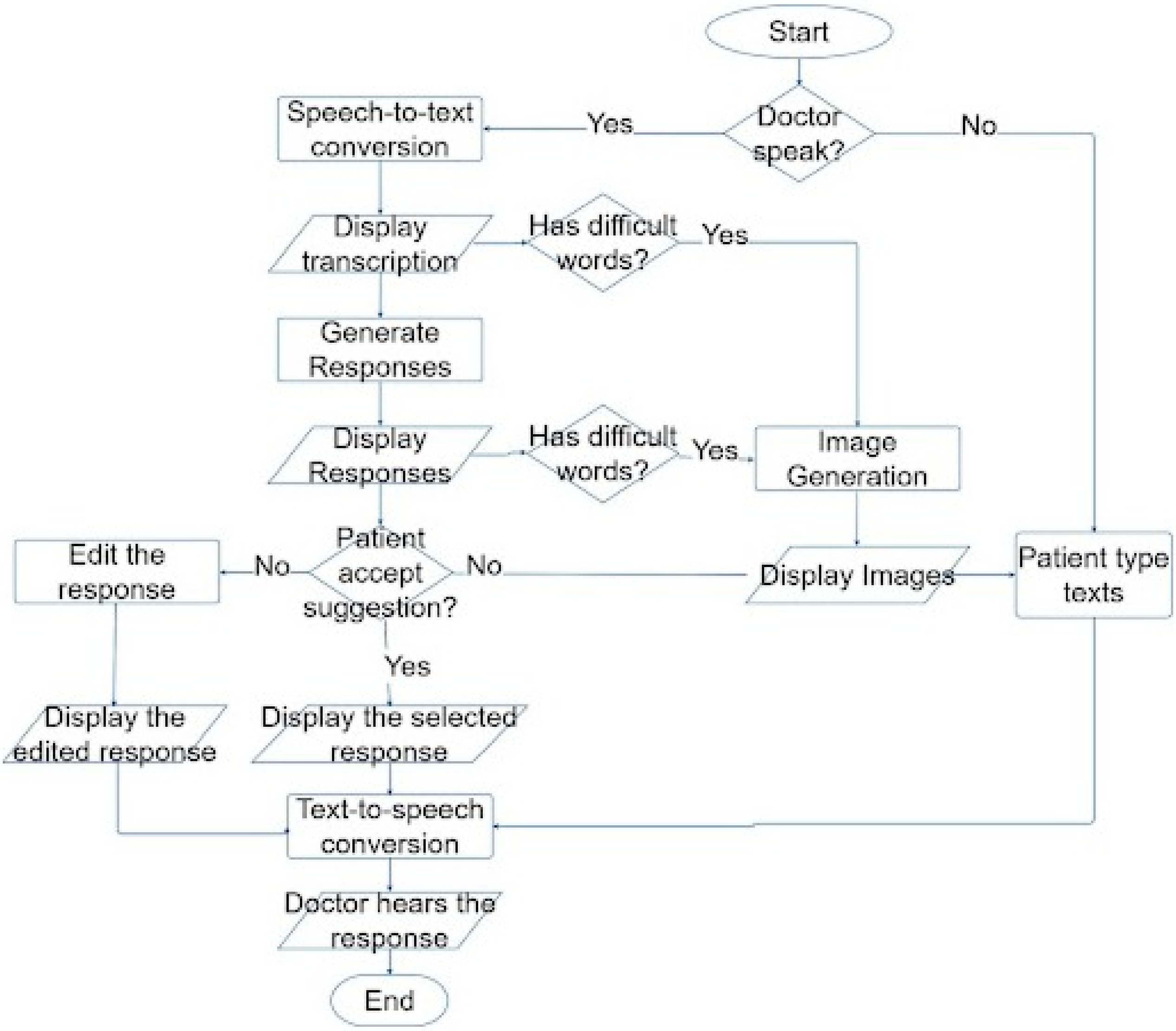


Figure 1.1: Operational Flow of the Developed Mobile application

The communication begins with the user selecting who initiates the conversation, either the healthcare provider or the DHH user. If the doctor initiates, their spoken input is captured by the application and immediately converted into readable text through real-time Speech-to-Text (STT) functionality. Based on the content of the doctor's message, the system generates a list of suggested response options. These suggestions are intended to assist the DHH user in formulating a quick and contextually appropriate reply. The user can either select a suitable suggestion, modify it according to their needs, or manually type a new response. For complex or unfamiliar vocabulary identified within the conversation, the application automatically provides corresponding visual aids, such as images and English translations, to enhance user comprehension.

Once the user finalizes their response, it is converted into spoken output using Text-to-Speech (TTS) technology, enabling the doctor to hear the message clearly. Conversely, if the DHH user chooses to initiate the conversation, they can directly type their message, which is then converted into speech for the healthcare provider. This interaction model supports fluid and accessible communication without reliance on lip-reading or sign language interpretation.

By providing a user-friendly, intelligent communication bridge, this mobile application empowers deaf individuals to manage their healthcare interactions with greater confidence and clarity. It supports independent communication while also promoting accessibility and understanding in medical environments [4].

1.1 Problem Statement

Effective communication is essential in all aspects of life, yet for individuals who are deaf or hard-of-hearing (DHH), it presents significant challenges, particularly in healthcare settings. DHH individuals often rely on sign language, lip-reading, or written communication, which can be limited and lead to misinterpretation or information gaps. In medical environments, the lack of accessible, real-time communication tools exacerbates these challenges, creating barriers to effective doctor-patient interactions. Healthcare professionals often lack the necessary training or resources to communicate effectively with DHH patients, leading to frustration, misunderstandings, and suboptimal care [5].

Despite advancements in assistive technologies, solutions that can bridge the communication gap in real-time medical interactions are still lacking. Current tools fail to address the full range of DHH needs, such as being context-aware, user-centered, and capable of supporting clear communication across various medical scenarios. This is especially critical in consultations, diagnoses, and follow-up visits, where the quality of communication directly affects the quality of care and health outcomes for DHH individuals. Thus, there is an urgent need for innovative, accessible, and practical communication solutions that can facilitate seamless interactions between DHH individuals and healthcare providers, ensuring that DHH patients receive equitable, effective, and empathetic care.

1.2 Research Questions

The purpose of this research project is to undertake a comprehensive exploration into the inquiries posed by the following research questions:

- **How to design and develop a user friendly communication application for deaf and hard of hearing individuals with hearing individuals?**

  The focus is on how to design and develop a communication application that facilitates effective interaction between deaf and hard of hearing individuals and hearing individuals. It involves exploring methods to create an accessible and user-friendly platform that addresses communication challenges faced by both groups. The question considers the integration of appropriate technologies and interface designs to support smooth, inclusive, and natural communication across different user needs.

- **How to reduce communication effort for deaf and hard-of-hearing individuals?**
  The focus is on finding methods to reduce the communication effort required by deaf and hard-of-hearing individuals during interactions. It involves exploring strategies, technologies, and design approaches that can make communication more efficient, intuitive, and less physically or mentally demanding. By minimizing the barriers in exchanging information, the research looks at ways to support smoother, faster, and more comfortable communication experiences.

- **How to improve the interaction speed to support real-time conversations with hearing individuals?**
  The focus is on identifying ways to improve interaction speed during conversations between deaf or hard-of-hearing individuals and hearing individuals. It involves exploring techniques and technologies that can minimize delays, streamline message exchange, and create a more natural, real-time conversational flow. The research question addresses how faster interaction can better support continuous, real-time communication without disrupting the engagement between participants.

- **How can the use of suggested responses and visual aids (such as images) improve the clarity of text-based communication and enhance social engagement for deaf or hard-of-hearing individuals in conversations with hearing individuals?**
  The focus is on how suggested responses and visual aids, such as images, can improve the clarity of text-based communication between deaf or hard-of-hearing individuals and hearing individuals. It examines how these features can make conversations more expressive, reduce misunderstandings, and support faster, more confident interactions. By strengthening the quality of communication, the use of these tools is expected to enhance social engagement and create a more natural conversation flow for deaf or hard-of-hearing users.

1.3 Goals and Objectives

1.3.1 Goals

This research aims to develop a system that improves communication between deaf/hard-of-hearing (DHH) individuals and non-hearing-impaired individuals [1]. The primary goals of the system are to enhance communication efficiency, accuracy, and user satisfaction. The focus is on improving human-computer interaction (HCI) to ensure that DHH individuals can communicate seamlessly with non-hearing-impaired individuals in various contexts, particularly in real-time interactions.

1.3.2 Objectives

1. **Evaluate Communication Efficiency**

   This research aims to assess how the placement, size, and visibility of text input and output fields within the application influence the communication efficiency of both Deaf and

Hard-of-Hearing (DHH) users and non-hearing-impaired users. By systematically analyzing task completion times, error rates, and user satisfaction during usability testing, the study seeks to identify optimal interface configurations that support faster, more accurate, and more comfortable communication experiences. Improving communication efficiency is critical to ensuring that the application facilitates real-time medical consultations effectively.

2. **Examine Design and Usability**

   This study also investigates the role of visual design elements - including color schemes, layout organization, button sizes, and iconography - in enhancing the usability and accessibility of the application for DHH users. Special emphasis is placed on evaluating how these design factors affect user engagement, perceived ease of use, and the overall accessibility of the system. By identifying the most effective design practices for DHH-friendly interfaces, the research aims to provide design recommendations that ensure the application is inclusive, intuitive, and culturally sensitive.

1.4 Research Approach

This research adopts a deductive research approach, beginning with established theories and frameworks in Human-Computer Interaction (HCI) and Deaf accessibility technologies and applying them to the design and evaluation of a mobile application aimed at facilitating communication between Deaf/Hard-of-Hearing (DHH) individuals and non-hearing-impaired individuals in a Sinhala language context. Through the deductive approach, the study tests specific assumptions regarding the effectiveness of integrating speech-to-text functionality, visual aids, and Sinhala script interfaces in improving communication efficiency and user satisfaction.

To achieve this objective, the research follows a User-Centered Design (UCD) methodology, which places DHH users at the center of the design and development process. UCD involves gathering user requirements, developing prototypes, conducting usability testing, and iteratively refining the application based on continuous user feedback. This ensures that the final product aligns closely with the real-world needs and expectations of its intended users.

A mixed-methods strategy is also employed, combining qualitative and quantitative research techniques throughout the design, development, and evaluation phases. Qualitative methods, such as interviews and observational studies, are used to explore user needs, identify usability issues, and understand emotional and experiential aspects of the interaction with the application. At the same time, quantitative methods, including the measurement of usability testing metrics such as task completion time, error rates, and structured questionnaire responses, are utilized to objectively assess the system's performance and user satisfaction.

The integration of deductive reasoning, UCD principles, and mixed-methods research ensures a comprehensive and robust understanding of the application's usability, communication effectiveness, and suitability for deployment in real-world healthcare settings involving DHH individuals.

1.5 Limitations, Scope and Assumptions

1.5.1 Scope

The scope of this research encompasses the development of a mobile application designed to facilitate communication between DHH individuals and non-hearing-impaired individuals.

- **Speech-to-Text (STT) Functionality**

  The Speech-to-Text (STT) functionality is a critical feature of the application, designed to capture spoken language from hearing individuals and convert it into written text in real time. This enables DHH users to read the spoken input directly on the application interface, thereby facilitating immediate and seamless communication. In the context of this research, implementing accurate and efficient STT was essential to ensure that DHH users could participate actively in conversations without the need for external interpretation services. The system's real-time transcription capabilities contribute significantly to reducing communication delays and misunderstandings.

- **Suggested Response Generation**

  To further support DHH users in responding during conversations, the application includes a Suggested Response Generation feature. This functionality presents users with a selection of predefined, contextually appropriate replies based on common conversational scenarios. Users can choose from these options, which are then vocalized using a Text-to-Speech (TTS) system to communicate back to the hearing individual. This method not only saves time compared to typing but also ensures that responses are grammatically correct and culturally appropriate. In this research, this feature was particularly important for users who may experience difficulties with fast text composition during dynamic interactions.

- **Image Generation**

  Recognizing that certain complex or unfamiliar words may pose challenges in text-based communication, the application also integrates an Image Generation functionality. When the system detects a difficult or uncommon word (especially medical or technical terms), it generates a corresponding visual representation to aid user understanding. This is particularly relevant in conversations involving specific terminology that DHH users might not immediately recognize. The visual support enhances comprehension, reduces confusion, and strengthens the overall communication process. Within the research, this feature was tested to evaluate its effectiveness in facilitating better understanding among DHH participants.

- **Sinhala Language Support**

  A major focus of the application was ensuring linguistic and cultural relevance by conducting all testing and development primarily in the Sinhala language. The Speech-to-Text, Text-to-Speech, Suggested Responses, and Image Generation features were customized and evaluated in Sinhala, addressing a significant gap in accessibility technologies for Sri Lankan DHH users. Sinhala support ensures that users interact with the system in their native language, promoting greater comfort, comprehension, and adoption. This focus on local language usability distinguishes this research from existing international solutions, many of which lack meaningful localization.

- **Customizable Interface**

  The application includes a customizable user interface, allowing users to modify appearance settings such as font size, background color, and text color based on their visual preferences and individual needs. Additionally, users can personalize certain functional aspects, such as enabling or disabling Suggested Responses or Image Generation, depending on their communication preferences. This customization ensures that the system is adaptable to a wide range of user requirements, including those with additional impairments like low vision. In the research, customization options were evaluated for their impact on user satisfaction and overall usability. These features aim to bridge communication gaps, promoting inclusivity and enhancing the interaction between DHH individuals and the hearing community.

These features aim to bridge communication gaps, promoting inclusivity and enhancing the interaction between DHH individuals and the hearing community.

1.5.2 Limitations

- **Environmental Variability**

  The accuracy and reliability of the speech-to-text (STT) system can be significantly impacted by environmental conditions, particularly background noise. Outdoor settings or busy clinical environments often introduce auditory disturbances that degrade the quality of voice capture and, consequently, the accuracy of transcription. In this research, while testing was conducted under controlled environments, it is recognized that real-world application scenarios may present challenges that could affect communication effectiveness.

- **Language Support**

  The scope of the application is currently restricted to supporting the Sinhala language for both input and output operations. While this focus was essential to address the linguistic needs of the primary user group within the Sri Lankan context, it inherently limits the system's applicability to users who communicate in other local languages (such as Tamil) or international languages. Expanding language support remains a consideration for future development phases.

- **Hardware Limitations**

  The performance of the application is influenced by the hardware capabilities of the mobile device on which it is installed. Factors such as microphone sensitivity, processing speed, and available memory can directly affect the accuracy and responsiveness of core functionalities like Speech-to-Text and Text-to-Speech. Devices with lower specifications may experience delays, reduced transcription accuracy, or application instability, thereby impacting user experience.

- **User Diversity**

  While the application is designed to cater to the DHH community broadly, this study does not differentiate participant outcomes based on factors such as varying literacy levels, educational backgrounds, or prior experience with mobile technology. These variables could influence user performance and satisfaction but were not specifically analyzed within the scope of this initial evaluation phase.

- **Live Medical Context Testing**

  Due to ethical, privacy, and logistical constraints, the system has not been extensively tested within real-time, high-pressure medical consultation environments during the initial stages of research. Simulated tasks and controlled usability testing were employed instead. Although the application demonstrated promising results under controlled conditions, its performance in actual clinical settings remains to be validated in future studies.

1.5.3 Assumption

Deaf Student Access to Mobile Devices

It is assumed that the target user group, namely deaf or hard-of-hearing (DHH) individuals, have access to smartphones or mobile devices with sufficient hardware capabilities (such as a functioning microphone, stable processing power, and adequate memory). The application is designed as a mobile-first solution, leveraging the ubiquity of smartphones among younger populations. If users lack access to compatible devices, the system's real-world applicability and inclusivity would be compromised. Therefore, ensuring device accessibility is critical for the application's success.

User Willingness

The study presumes that DHH individuals are open to adopting technological interventions to support their communication needs, particularly in sensitive contexts such as medical consultations. User willingness is fundamental because technological acceptance determines real-world usage. If users are resistant to using new technology -due to distrust, cultural barriers, or usability concerns - the system's benefits would not be fully realized. Hence, this research assumes a positive attitude toward assistive mobile technology among the target population.

Data Privacy Compliance

It is assumed that the application complies with all relevant data protection and privacy regulations, such as GDPR or local data protection laws. Since the system handles personal and potentially sensitive health information during consultations, ensuring data confidentiality and security is ethically mandatory. This assumption acknowledges that safeguarding user trust through secure handling of data is integral to user acceptance, particularly in healthcare settings where privacy breaches could lead to significant harm.

User Proficiency in Written Language

The application relies on users being able to understand written Sinhala to interact effectively with the system. It is assumed that participants possess at least a moderate level of literacy, enabling them to read transcribed speech and select or type responses when necessary. Without sufficient literacy, the system's text-based features, such as reading generated responses or interpreting real-time STT outputs, would become inaccessible, thereby reducing usability and limiting communication success.

User Training and Familiarity

It is assumed that users either have prior familiarity with mobile applications or can quickly learn to operate the proposed system after minimal instruction. The application was designed to be user-friendly; however, a basic understanding of touchscreen interaction, navigation through menus, and reading visual prompts is necessary. If users lack technological literacy, additional training efforts would be required, which were not within the scope of this study. Therefore, the assumption simplifies testing and deployment, focusing on usability evaluation rather than basic digital literacy training.

Environmental Suitability

The application's performance assumes the presence of conducive environmental conditions in the healthcare setting where it is employed. Ideal conditions include adequate lighting (for visual interaction with the screen), minimal background noise (for accurate speech recognition), and stable internet connectivity (for loading additional resources such as images for complex terms). If the real-world environment is significantly more challenging, system effectiveness , particularly features like Speech-to-Text or image generation ,could be impaired. Thus, this assumption underpins the expectation of reasonably supportive external factors during application usage.

1.6 Contribution

Development of a Real-Time Communication Mobile Application for Deaf individuals in Healthcare Settings

This research introduces a mobile application designed to bridge the communication gap between deaf individuals and healthcare professionals during medical consultations. The application enables seamless interaction by converting spoken language into text and vice versa, allowing for real-time communication without the need for sign language interpreters or lip-reading. This innovation enhances accessibility and autonomy for deaf individuals in healthcare environments.

Evaluation of User Experience and Effectiveness in Enhancing Healthcare Communication

This study evaluates the usability and effectiveness of the developed mobile application in real-world healthcare scenarios. Through user testing and feedback from deaf individuals , the research assesses the application's impact on communication clarity, patient satisfaction, and overall healthcare experience. The findings provide valuable insights into the practical benefits and potential challenges of implementing such technology in medical settings, informing future developments and applications in the field of assistive healthcare technologies.

Chapter 2

Background

2.1 Communication Challenges Faced by Deaf individuals in Healthcare

Different levels of auditory impairment, from partial hearing loss to total deafness, are referred to as deafness and hard of hearing (DHH). According to decibel (dB) thresholds, the World Health Organization (WHO) divides hearing loss into four categories [6] : mild (26–40 dB), moderate (41–60 dB), severe (61–80 dB), and profound (equal to or greater than 81 dB). While people with residual hearing may utilize assistive technologies (such as cochlear implants or hearing aids) or speechreading, profoundly deaf people typically use sign language as their primary form of communication.

Due to a lack of access to specialized education, many DHH people in Sri Lanka, where Sinhala is the primary language, struggle with literacy, which makes communication more difficult in crucial contexts like healthcare.

When seeking to obtain healthcare services, Deaf and hard-of-hearing (DHH) people face substantial communication challenges, which frequently result in worse health outcomes. The lack of trained sign language interpreters in medical contexts is one of the main problems. Many clinics and hospitals lack qualified interpreters who are fluent in local sign languages, especially in developing nations like Sri Lanka. Even when they are accessible, interpreters might not be well-versed in medical jargon, which could lead to translation errors. Because of this, a lot of DHH patients are forced to communicate with family members, which affects privacy and could result in miscommunications about private medical information.

Another major barrier is health literacy. Due to educational gaps, many DHH people have poorer reading levels, which makes it challenging for them to comprehend written medical instructions or fill out health documents. This problem is further compounded by complicated medical jargon, since words like ”antibiotics” might not be commonly used. Although some medical professionals try to communicate with DHH patients in writing, this approach is frequently unsuccessful because written correspondence takes a lot of time, and patients’ literacy abilities vary.

Barriers related to culture and attitude make these difficulties worse. Numerous medical professionals are ignorant of Deaf culture and communication requirements, and they occasionally assume the wrong things about cognitive ability based solely on hearing status. Even for highly proficient lip readers, lip reading accuracy is only 30–40%, and some clinicians overestimate its efficacy. These misconceptions may result in incorrect diagnoses, insufficient treatment, and patient annoyance.
These multifaceted problems show how urgently innovative approaches that simultaneously address language, technological, and cultural limitations are needed.

2.2 The Need for Assistive Technology in Sinhala Healthcare Contexts

With the increasing availability of mobile technology, there is a growing opportunity to bridge communication gaps through intelligent applications. However, most existing apps focus on English or international use cases, offering little to no support for local languages like Sinhala. Moreover, the needs of deaf individuals who may be more comfortable with Sinhala and visual learning methods are often overlooked in mainstream app development.

This research addresses that gap by designing and developing a mobile app that helps deaf individuals communicate with medical professionals in Sinhala, incorporating speech-to-text, Sinhala script support, enhanced response generation, and visual aids for difficult terms. This solution is not only linguistically appropriate but also culturally sensitive and context-aware.

2.3 Key Features to Support Communication

The application integrates several core features specifically designed to enhance communication between Deaf and Hard-of-Hearing (DHH) individuals and healthcare providers during medical consultations. These features were informed by findings from usability studies, interviews with DHH participants, and established best practices in Human-Computer Interaction (HCI). The overall aim is to increase communication efficiency, user comfort, and user autonomy in clinical settings.

- **Real-time Speech-to-Text Conversion**
  The system provides real-time conversion of spoken Sinhala language into readable text displayed on the user's mobile device. This feature enables deaf individuals to immediately understand spoken communication from healthcare providers without relying on interpreters or written notes. The STT functionality supports spontaneous, dynamic conversations, helping to create a more natural consultation experience and minimizing communication delays.

- **Response Suggestions**
  The application offers intelligent, contextually relevant reply suggestions based on the ongoing conversation. DHH users can select from these predefined responses, which are subsequently converted into spoken output using Text-to-Speech (TTS) technology. This feature helps users respond more quickly and easily, enabling them to participate in real-time medical discussions without the need to manually type full sentences during the consultation.

- **Visual Aids for Difficult Words**
  To support the comprehension of complex or unfamiliar terms often encountered in medical conversations, the application automatically identifies difficult Sinhala words. It then provides English translations and displays relevant visual representations (images) to assist understanding. This multimodal support enhances the user's ability to grasp specialized vocabulary, especially critical in healthcare environments where accurate understanding can impact medical decision-making.

- **Sinhala Script Interface**
  The entire user interface of the application is developed in Sinhala script, ensuring cultural and linguistic relevance for local users. By prioritizing the native language and writing system of the intended user group, the application maximizes accessibility and reduces language barriers, promoting more inclusive healthcare communication.

These features were informed by usability studies, interviews with deaf individuals , and best practices in human-computer interaction (HCI), with the goal of improving communication efficiency, user comfort, and autonomy during medical interactions.

Chapter 3

Literature Review

The purpose of this literature review is to examine relevant research conducted by others in the field and to identify existing research gaps. We reviewed studies primarily through Google Scholar, focusing on those that address communication needs and accessibility for deaf and hard-of-hearing individuals. Our analysis aims to clarify how current technologies support, or fall short in supporting, the unique communication needs of this community. This review covers prior research on text-to-speech, speech-to-text, and image-processing technologies, specifically examining studies that target the needs of deaf and hard-of-hearing individuals.

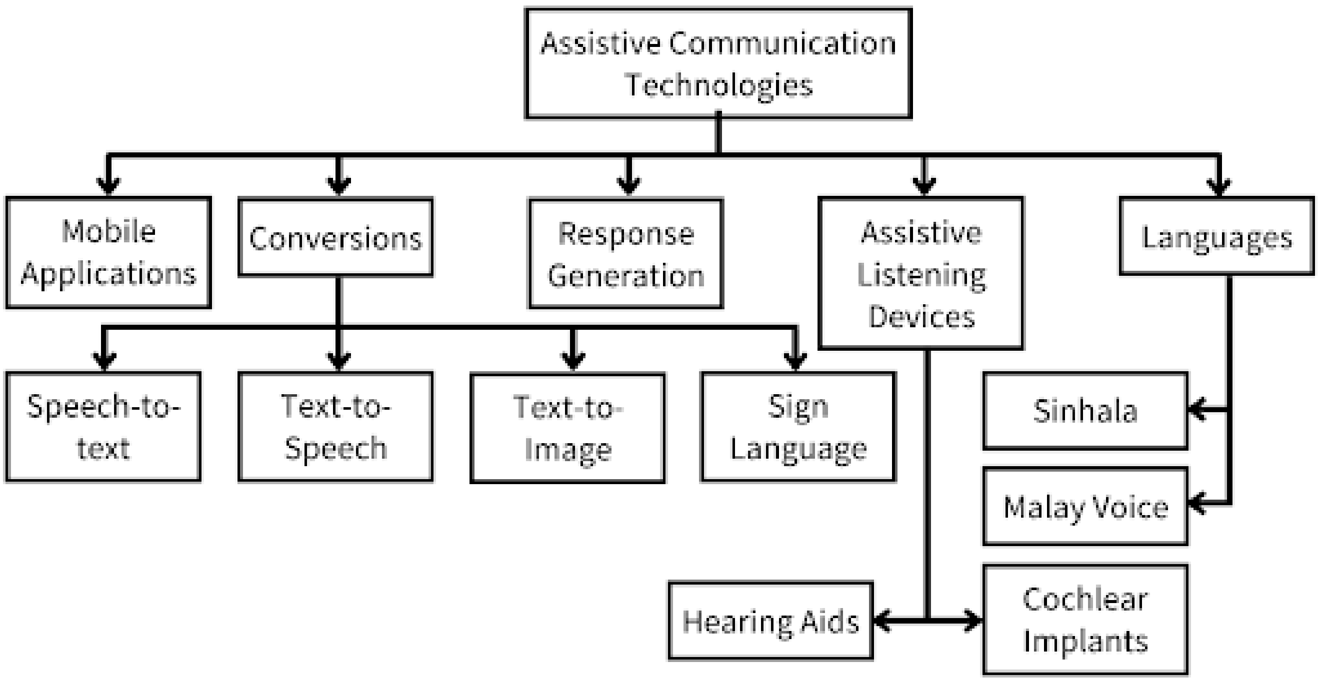


Figure 3.1: Taxonomy for use of Assistive Communication Technologies

3.1 Understanding the nature of the DHH

Hearing loss leads to various communication challenges for people. In the deaf community, the word deaf is spelt in two ways. The small "d" deaf represents a person's level of hearing through audiology. It is not associated with the other members of the deaf community, whereas the capital

"D" Deaf indicates the culturally Deaf people who use sign language for communication [7]. The Deaf community is not a monolithic group; it has a diversity of groups, which are as follows [8]:

Group	**Description**
Hard-of-hearing people	They are neither fully deaf nor fully hearing. They can obtain some useful linguistic information from speech.
Culturally deaf people	They might belong to deaf families and use sign language as the primary source of communication. Their speech clarity may be disrupted.
Congenital deaf people	They are deaf by birth or become deaf before they learn to talk and are not affiliated with Deaf culture. They might or might not use sign language-based communication.
Orally educated deaf people	They have been deaf in their childhood but developed speaking skills.
Late-deafened adults	They have had the opportunity to adjust their communication techniques as a result of their progressive hearing loss.

Table 3.1: Diversity within the Deaf community

Hearing disorders have significant negative effects on the social and professional lives of individuals, particularly older adults. In children, hearing impairment is associated with persistent behavioral challenges and delays in language development. Studies have shown that children using hearing aids perform significantly better than those who do not, as hearing aids reduce the cognitive load required for speech processing, thus mitigating mental fatigue [9]. This highlights the potential of technology-based interventions in enhancing the lives of individuals with hearing impairments.

3.2 Assistive technologies used by DHH for communication

Various technologies have been developed to address communication challenges faced by individuals with hearing impairments. These technologies aim to improve accessibility and facilitate better communication in environments primarily designed for hearing individuals. The role of Assistive Listening Devices (ALDs) in supporting individuals with hearing impairments, including hearing aids, cochlear implants, and auditory therapy, is important. These devices help improve communication clarity in various environments, including educational settings. ALDs are key for creating inclusive environments, improving academic performance, and promoting social integration [10].

3.2.1 Hearing Aids

Hearing aids, as one of the most widely used devices for hearing impairment, offer substantial benefits but also face several challenges [11]. The distinction between analog and digital hearing aids is crucial in understanding their function. Analog hearing aids amplify all sounds indiscriminately, which can be problematic in noisy environments. In contrast, digital hearing aids convert sound into digital signals, offering noise reduction and more precise amplification. Despite these advancements, digital hearing aids still face challenges in fluctuating noise environments, and their effectiveness can be limited in complex listening situations. Additionally, they do not fully restore hearing, especially for individuals with profound hearing loss, as they cannot address the auditory system's fundamental limitations.

3.2.2 Cochlear Implants

Cochlear implants represent a more invasive solution to hearing impairment, offering substantial benefits for individuals with severe to profound sensorineural hearing loss. By directly stimulating the auditory nerve, cochlear implants bypass the damaged cochlea, offering a form of auditory perception that hearing aids cannot achieve [12]. While they provide better auditory perception than hearing aids, cochlear implants still face limitations, such as providing mechanical or distorted sound quality. The implantation process is surgical and carries risks like infection or device malfunction. Moreover, post-implantation rehabilitation is often required, and success rates vary depending on factors like age and the degree of prior hearing loss. These implants do not replicate natural hearing and may not be fully effective for all users.

3.2.3 Digital Sign Language Interpreters

Digital sign language interpreters, which utilize motion sensors and video mapping technologies, have the potential to bridge communication gaps between deaf individuals and the hearing community [13]. However, these systems struggle to capture the full range of communication in sign language, particularly non-manual signals like facial expressions, which are critical for accurate translation. Additionally, the quality of these systems depends on the technology used, and errors in translation can lead to misunderstandings. These systems cannot replace human interpreters who provide contextual and cultural understanding, limiting their overall effectiveness in real-world communication scenarios.

3.2.4 Assistive Technology for Higher Education individuals with Disabilities

A study looks at the assistive technologies used by individuals with hearing impairments in higher education, focusing on the tools that help them in their academic and personal lives. It also highlights the barriers they face in accessing and using these technologies [14].

The key findings show that individuals use several types of assistive technologies. Personal amplification devices, like hearing aids and cochlear implants, help enhance sound for those with hearing loss, though some individuals find them uncomfortable or worry about their visibility. Environmental control systems, such as induction loops or infrared systems, are commonly used in classrooms. These systems connect to hearing aids to provide clearer audio in noisy environments. Transcription and captioning tools are also essential. Text-to-speech and voice recognition software, like Dragon Naturally Speaking, convert spoken words into text, allowing individuals to follow along in real-time. Automatic subtitles and closed captions on videos and lectures are also critical for individuals with profound hearing loss. Video Call solutions, such as ntouch, help make video calls more accessible by providing real-time captioning.

The study also identifies several challenges faced by individuals with hearing impairments. One major issue is the lack of availability and support for assistive devices on campus, often due to limited funding or a lack of awareness from university staff. Social stigma is another challenge, as some individuals feel self-conscious about using visible devices like hearing aids, which can prevent them from using them in social or classroom settings. Additionally, many institutions lack formal policies and support for assistive technology, leading to inconsistent access across campuses.

ALDs are crucial for improving communication but remain limited by issues like comfort, visibility, and reliance on external systems. More development is needed on wearable, discreet, and comfortable solutions, as well as enhancing the integration of ALDs with other assistive tools to create a more seamless experience.

3.3 Text-to-speech and speech-to-text conversions

Speech-to-Text (STT) and Text-to-Speech (TTS) technologies support the educational needs of deaf and hard-of-hearing individuals by bridging the gap between spoken and written language. At the Model Secondary School for the Deaf (MSSD), these tools helped 14-year-old Mateo, who struggled with literacy. Reading at a pre-kindergarten level and not fluent in English or American Sign Language (ASL), he used TTS and STT to improve his reading and writing skills in both Spanish and English, enhancing his educational engagement. These technologies helped Mateo overcome challenges in literacy and allowed him to engage more fully in his education [15]. While demonstrating the potential of TTS and STT in supporting literacy development, the approach overlooks some limitations, such as challenges in translating complex or idiomatic expressions. Additionally, the effectiveness of these technologies in diverse learning environments may vary, requiring further exploration of personalized solutions for different users.

The "Deaf Chat" application utilizes advanced speaker recognition technology to improve communication for individuals with hearing impairments. It uses multiple-speaker classification to differentiate and transcribe speech in multi-speaker environments, which is a significant advantage over traditional hearing aids. The app leverages IBM Watson's Speech-to-Text API, with speaker diarization features that help identify and transcribe different speakers in group settings. The use of AI-driven speech recognition technologies shows promise in real-time communication, but accent recognition remains a significant challenge, limiting the app's effectiveness for diverse user groups. Further research could focus on improving the app's adaptability to different accents and dialects, as well as enhancing its performance in noisy environments.

A study used the SOVORO mobile application, which converts speech to text with a slight delay, as the target application to improve communication and promote social inclusion for DHH [16]. It includes features such as voice caption generation, storage, and playback. The approach aimed to overcome participant limitations while effectively identifying universal usability issues with

Figure 3.2: A Chat Activity showing the processed voices of four distinct speakers.

expert input. The app's functionality was tested with 13 hearing-impaired participants in an educational setting. While the app was appreciated for its real-time functionality, issues related to accuracy in noisy environments and dialect-heavy speech were highlighted. Expert evaluators also noted that the app lacks flexibility and error recovery options, which could undermine user trust in critical situations. The app's strengths lie in its real-time functionality and potential for use in educational contexts. However, it faces challenges in noisy or complex environments, where its performance could be significantly hindered. The lack of error recovery features is a critical limitation, suggesting the need for more robust systems that can handle misinterpretations and maintain user trust.

A Speech-to-Text system designed for individuals with permanent hearing loss, specifically targeting Kannada speakers processes speech samples using Mel Frequency Cepstral Coefficients (MFCCs), Discrete Wavelet Transforms (DWT), and Vector Quantization (VQ) techniques. The

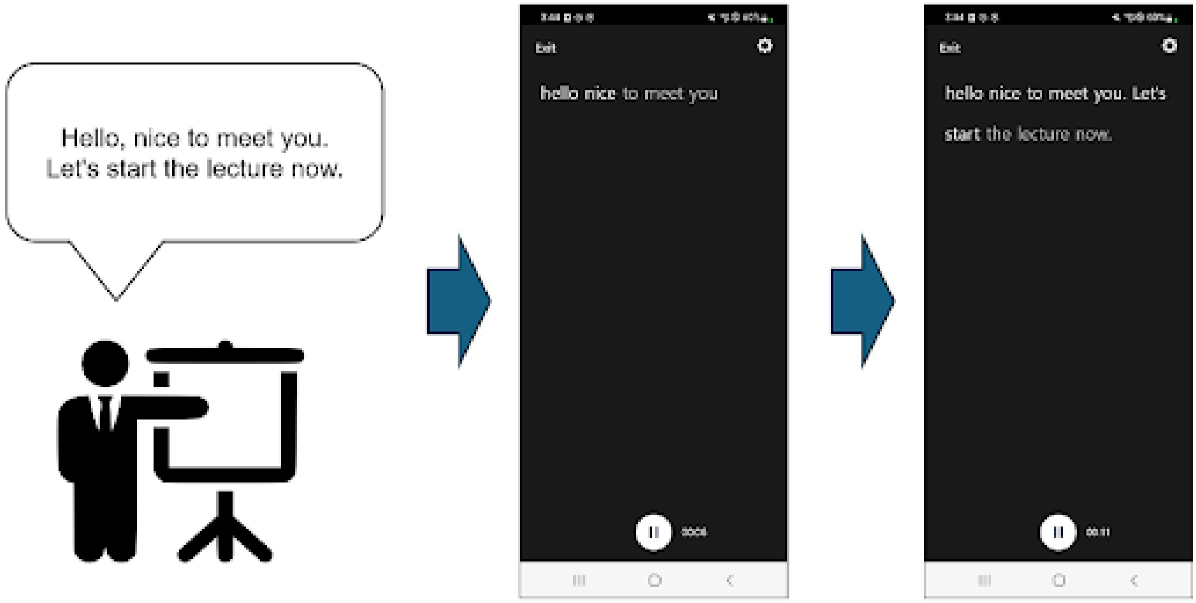


Figure 3.3: SOVORO Application

results indicate high recognition accuracy when using data from trained users, demonstrating the system's potential for continuous speech recognition and expansion to other languages [17]. The system is effective for isolated word recognition in controlled conditions but lacks the ability to handle continuous speech. Moreover, its reliance on a specific language limits its scalability. Future research should focus on expanding its capabilities to handle real-world speech variability and multilingual support.

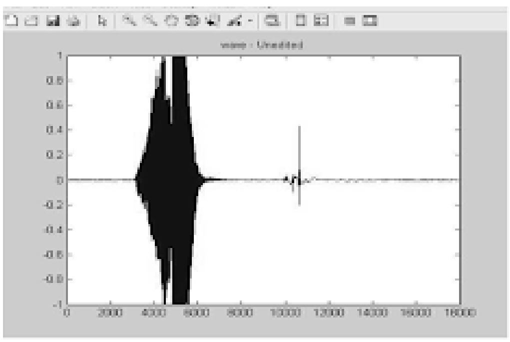


Figure 3.4: Wave of speech from database and test signal for sample one (ondu)

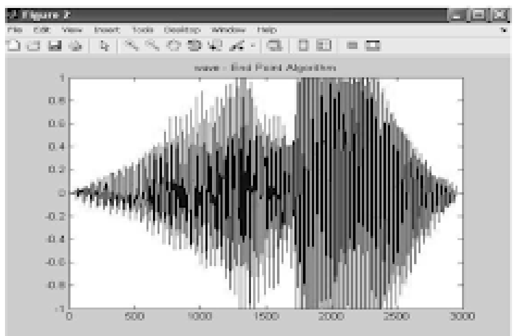


Figure 3.5: Test signal for sample one (ondu) after applying end point

Comuno is a mobile application developed to support individuals with hearing and speech impairments. It integrates Text-to-Speech (TTS) and Speech-to-Text (STT) technologies, enabling users to convert spoken language to text and vice versa. The app also stores conversations for future reference, facilitating communication and learning. The developers built Comuno using an iterative development model, leveraging technologies like the Ionic Framework, AngularJS, and Firebase for real-time data management and cross-platform compatibility (Android and iOS). The user interface is simple and intuitive, featuring icons for key functions: Listen (speech-to-text), Speak (text-to-speech), and Conversation (viewing saved interactions). User evaluations indicated that while the app effectively enhances communication, improvements are needed in real-time usability and accuracy [18]. The app's reliance on mobile platforms and limited user testing suggests that further development is required to optimize its functionality in diverse real-world contexts, such as classrooms or group discussions.

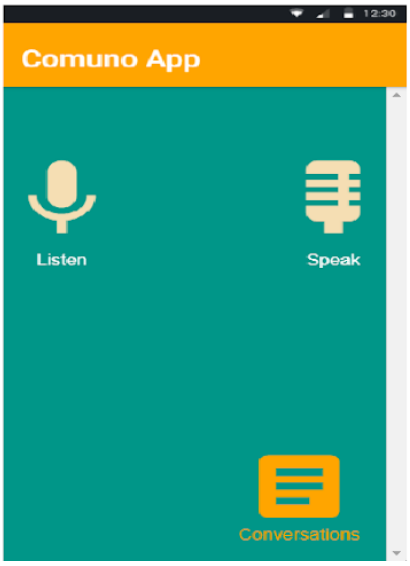


Figure 3.6: Comuno App Home Page Android view

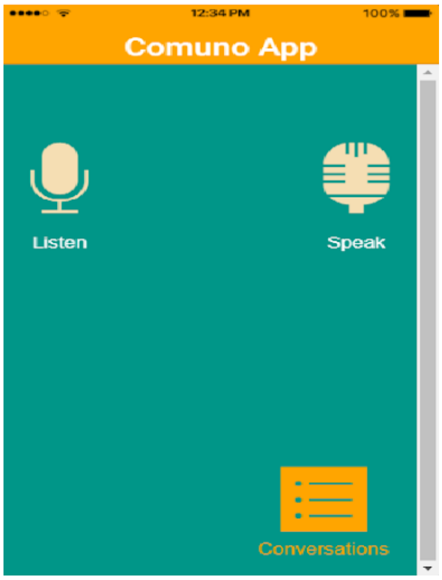


Figure 3.7: Comuno App Home Page IOS view

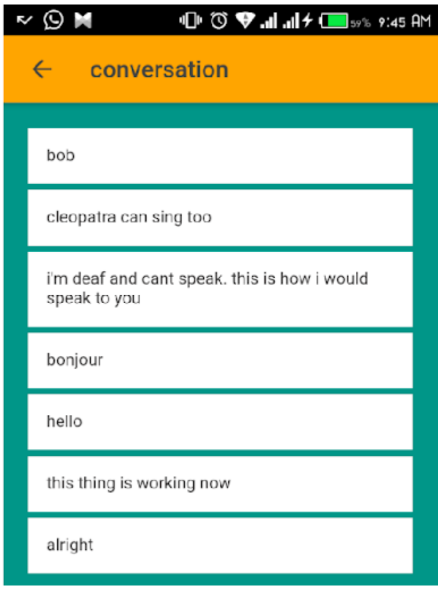


Figure 3.8: Android Interface Conversation Page with saved messages

Automated Speech Recognition (ASR) systems have demonstrated significant improvements in accuracy across devices such as smartphones. An evaluation assessed the performance of four ASR applications—AVA, Earfy, Live Transcribe, and Speechy—using Dutch audiological speech tests to compare the performance of normal hearing and hearing-impaired listeners [19]. The tests included the CNC test for phoneme recognition, the Digits-in-Noise (DIN) test, and the Plomp test for sentence recognition in noise. All apps achieved a minimum of 50% accuracy on the CNC test at 65 dB SPL, with performance improving at higher intensities. AVA and Live Transcribe achieved the lowest (best) signal-to-noise ratio (+8 dB) on the DIN test, while error rates were lower for English (19-34%) compared to Dutch (25-66%). The study provides valuable insights into the performance of ASR apps, but it overlooks certain contextual factors, such as the need for continuous speech recognition or the app's adaptability to different languages and accents. Further advancements in ASR could involve incorporating more robust noise-filtering technologies and real-time performance optimization.

Figure 3.9: Set-up of the smartphone in front of the speaker.

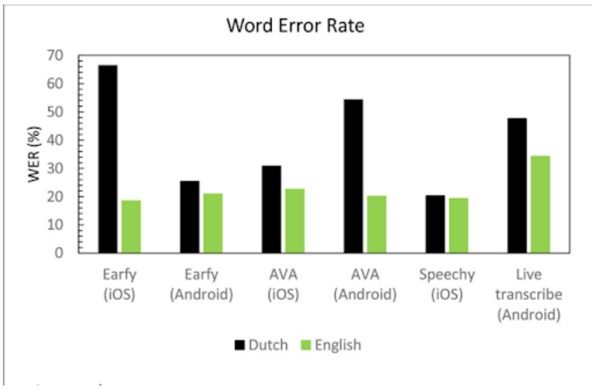


Figure 3.10: Word error rate in percentage of the dialogue in English and Dutch for the different apps.

3.4 Response Generation

Effective communication between deaf and hearing individuals in medical contexts requires not only accurate speech-to-text and text-to-speech conversion but also intelligent response generation to facilitate smooth, bidirectional dialogue. AI-assisted response generation introduces a critical layer of interactivity by suggesting contextually appropriate replies. Recent advances in natural language processing (NLP) have expanded the capabilities of assistive technologies for deaf and hard-of-hearing (DHH) individuals, moving beyond unidirectional solutions (e.g., speech-to-text) to interactive systems with AI-generated response suggestions. The integration of response generation enables more natural, bidirectional dialogue. Modern systems leverage large language models (LLMs) with chain-of-thought (CoT) prompting to enable more natural, context-aware dialogues [20].

The application of CoT prompting - where models generate intermediate reasoning steps before final responses - has shown particular promise in healthcare communication. This approach addresses two critical challenges in medical response generation: (1) ensuring clinically appropriate outputs, and (2) maintaining transparency in the AI's reasoning process [21]. Research demonstrates that when LLMs are prompted to ”think step by step” [22], they produce more accurate and interpretable responses compared to direct answer generation - a crucial factor for medical applications where error prevention is paramount [23].

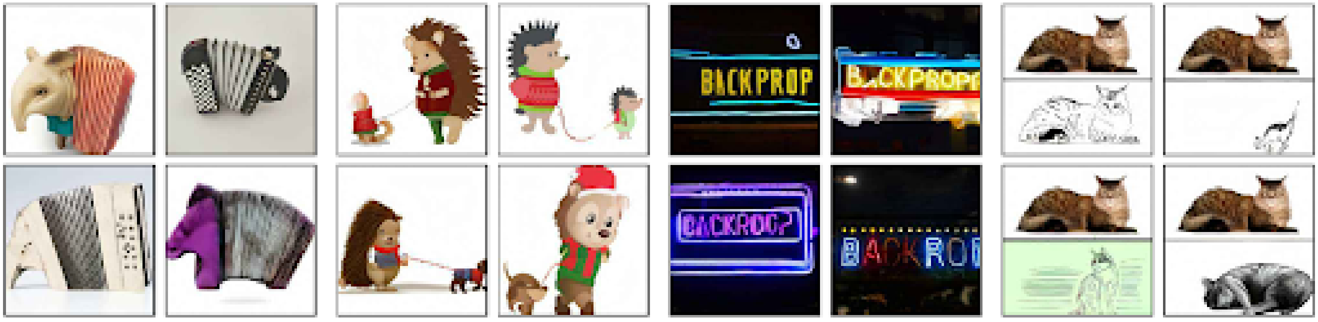
Large language models (LLMs) powering modern AI systems such as Google's Gemini, OpenAI's GPT models demonstrate significant potential in generating context-aware responses for DHH users [24]. A recent research demonstrates the potential of LLM-powered response generation to enhance accessibility for DHH individuals in educational contexts [25]. Their system, which generates context-aware Visual and Emotion questions based on real DHH learner data, highlights key considerations for medical communication tools: (1) the importance of editable, multimodal outputs (text/visual cues) to accommodate diverse language proficiencies, and (2) the need to integrate domain-specific data (e.g., medical dialogue patterns) to ensure clinical appropriateness. While their focus was educational, their findings align with challenges in healthcare communication—particularly the balance between automation and user agency, as DHH participants preferred modifying AI-generated responses rather than accepting them verbatim. This underscores the value of adaptive response generation in clinical settings, where accuracy and personalization are critical.

3.5 Image Generation

For DHH individuals, visual and textual cues play a central role in understanding stories, as they rely more on these elements than on spoken language. The connection between visual representations and language is still essential, but it is experienced differently from hearing individuals [26]. Text-to-image generative models, including auto-regressive and diffusion-based approaches, have seen substantial advancements in recent years, enabling the generation of high-fidelity images from text descriptions.

A study was conducted on a two-stage training process that enhances text-to-image generation by addressing challenges in image tokenization and model scalability [27]. In the first stage, a discrete variational autoencoder (dVAE) is used to compress images into tokens, thereby reducing memory usage while preserving visual quality. Unlike pixel-level tokenization, which may overlook key features, the dVAE method optimizes tokenization by employing Gumbel-Softmax relaxation and the straight-through estimator to refine discrete distributions over image tokens.

In the second stage, a transformer model processes both text and image tokens together, capturing intricate relationships between the two modalities. Custom attention masks are applied to improve token handling, while mixed-precision training facilitates the handling of large-scale datasets such as Conceptual Captions.
For image generation, the model generates multiple candidate images from text and selects the most appropriate one through a contrastive reranking process, which enhances both accuracy and realism. However, this method may limit accessibility due to its high computational demands. Despite the model's strong performance, its substantial computational requirements and environmental impact pose significant challenges for sustainable deployment.



(a) a tapir made of accordion. a tapir with the texture of an accordion.

(b) an illustration of a baby hedgehog in a christmas sweater walking a dog

(c) a neon sign that reads "backprop". a neon sign that reads "backprop". backprop neon sign

(d) the exact same cat on the top as a sketch on the bottom

Figure 3.11: The model demonstrates varying reliability in combining distinct concepts, creating anthropomorphized animals, rendering text, and performing image-to-image translation.

3.6 Sinhala Language

Our application is designed specifically for a user group composed primarily of Sinhala-speaking individuals who are deaf or hard of hearing. For this demographic, Sinhala is not only the native language but often the only language they understand and use in daily life. Due to limited exposure to English or other languages, using Sinhala ensures effective communication, improved comprehension, and a sense of cultural familiarity. Choosing Sinhala is therefore a deliberate and user-centered decision grounded in both practicality and inclusivity, ensuring the application remains accessible to the community it aims to serve.

Research supports this decision by highlighting the critical role of native language in assistive technology design. Samaradivakara et al. introduced an AR-based captioning interface developed specifically for Sinhala-speaking deaf and hard-of-hearing students. Their findings demonstrated significantly higher engagement and understanding when captions were delivered in Sinhala, confirming the impact of native-language content on accessibility and learning outcomes [28]. Similarly, Perera et al. developed an intelligent assistant using Sinhala for interaction, showing that localized interfaces empower users and enhance digital communication for hearing-impaired individuals [29].

The linguistic divide between spoken Sinhala and Sri Lankan Sign Language (SSL) is another important consideration. As noted by Nethsinghe and Udugama, there exists a substantial disconnect between Sinhala and SSL, with limited standardized mapping between the two. This gap underscores the need for applications that can translate and interpret between these modalities, further justifying Sinhala's inclusion in apps meant for this user group [30]. Moreover, studies by Gamage et al. and Gunasekara et al. present compelling technical frameworks that enable Sinhala sign language translation through avatars and AI-driven systems, demonstrating the growing feasibility and innovation in Sinhala-language accessibility tools [31], [32].

3.7 Malay voice

Although Sinhala is the primary textual and sign-based language in our application, the integration of a text-to-speech (TTS) system was limited by the unavailability of high-quality Sinhala voice synthesis engines. Most Sinhala TTS tools currently available are either experimental or insufficient for real-time usage, as highlighted by Dias and Jayasena [33]. While Divehi is linguistically closer to Sinhala, it also lacks usable TTS support or robust datasets for speech applications, as noted by Bestgen [34].

Given these limitations, Malay was selected as the TTS output language due to its phonetic and rhythmic similarity to Sinhala and its broad support across major platforms like Google Cloud. Prior research in speech modeling for low-resource languages supports the adaptation of phonetically similar TTS voices when native-language models are absent. Stüker [35] and Jayawardhana & Aponso [33] describe this method as a viable interim solution in multilingual and accessibility systems. Additionally, Coperahewa [36] points out historical and cultural interactions between Sinhala and Malay communities, which have contributed to shared phonological elements—reinforcing the functional compatibility of this choice.

Thus, the selection of Malay TTS provides a practical and intelligible voice output for Sinhala text, making the application more accessible to partially hearing-impaired users, and enhancing auditory feedback and user interaction in the absence of native voice resources.

3.8 Research Gap

Many assistive devices facilitate communication within the deaf community, with a predominant focus on sign language conversion. Sign languages exhibit significant variability, with differences in hand shapes, motion profiles, and the positioning of the hands, face, and other body parts contributing to the distinctiveness of each sign. This complexity makes visual sign language recognition a challenging area of research in computer vision [37], [38]. However, as sign language varies across regions, devices supporting languages such as American Sign Language present accessibility challenges for the Sri Lankan Sinhala deaf community. This issue is compounded by high levels of illiteracy within the community, particularly affecting individuals who struggle with more complex vocabulary, further hindering effective communication [39]. Visual aids, including images, have been identified as a potential solution to address these challenges [40].

Additionally, Automatic Speech Recognition (ASR) systems have primarily been developed for global languages like English, resulting in limited accessibility for Sinhala speakers. Therefore, the need for a Sinhala-specific ASR system is evident. Furthermore, the integration of next-word prediction systems has been recognized as a means to enhance both accessibility and usability for this population [41].

3.9 Technologies

Popular Speech-to-Text (STT) solutions used in related works, such as Google Speech-to-Text, IBM Watson STT, and Microsoft Azure STT, are commonly used for transcribing spoken language into text. However, these systems face challenges when transcribing Sinhala, particularly in noisy environments or with diverse accents and dialects. Their performance in handling Sinhala audio is often less accurate compared to other languages. To address this, we selected Google Speech-to-Text API for its ease of integration and support for real-time transcription, while recognizing the need to continually evaluate its performance with Sinhala inputs.

For generating spoken output, we utilize the Google Text-to-Speech (TTS) API, which provides clear and natural voice synthesis. Although Sinhala voice support is limited, the available voices (such as Malay) are used as alternatives to maintain communication effectiveness.

In terms of response generation, we use the Gemini API, which allows for dynamic and context-aware answers based on the user's input. This enhances interactivity and ensures that the system can adapt responses based on the communication context, improving the overall user experience.

For image generation, we rely on a combination of the Google Translate API and Wikimedia API. The Translate API is used to convert difficult Sinhala or English terms into simpler alternatives or definitions, which are then used to fetch relevant and illustrative images from Wikimedia. This supports better visual understanding for users with hearing impairments, especially in medical contexts where clarity is essential.

By combining these technologies—Google APIs for STT and TTS, Gemini API for intelligent responses, and Google Translate with Wikimedia for contextual image generation—we create a seamless, accessible, and intelligent communication platform tailored for deaf individuals in medical settings.

Chapter 4

Methodology

The research was conducted utilizing Design Science Research (DSR) methodology. The framework suggests a six-step iterative approach, beginning with problem identification and motivation and ending with evaluation and communication [42].

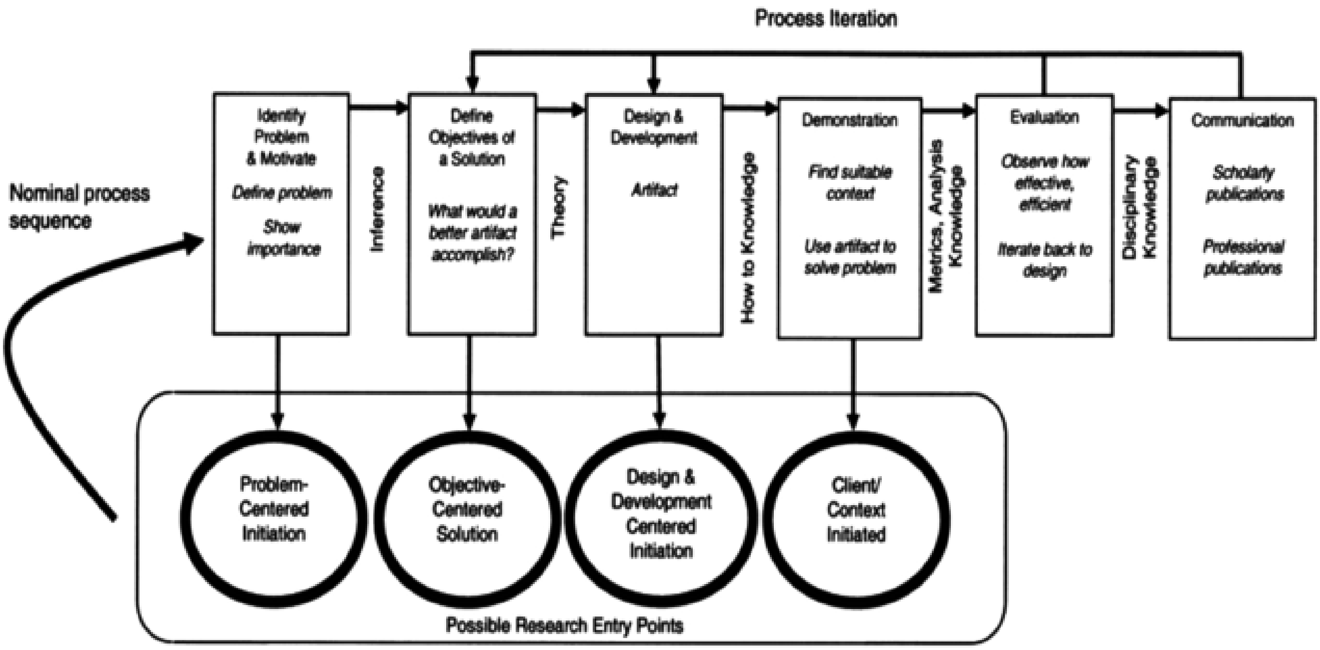


Figure 4.1: Design Science Research Methodology

DSR is a well-established and widely recognized methodology within the field of Information Systems (IS) research, particularly suited for studies focused on the development and evaluation of novel technical artifacts that address significant real-world problems. Given that the objective of this research is to develop a mobile application to address a critical gap in medical communication for deaf and hard-of-hearing individuals within Sinhala-speaking contexts, the selection of DSR as the guiding methodology is particularly appropriate. Furthermore, DSR facilitates the iterative creation, testing, and refinement of artifacts while simultaneously generating insights that contribute to both academic knowledge and practical application.

4.1 Problem Identification and Motivation

The initial step of the first phase of our research involved problem identification and motivation. A comprehensive literature study was carried out to evaluate existing solutions for DHH communication in healthcare, indicating a severe shortage in Sinhala-language assistive aids for medical consultations. The key issues found were,

- A lack of real-time bidirectional contact between doctors and DHH (Deaf and Hard of Hearing) patients.
- There is limited speech-to-text (STT) and text-to-speech (TTS) support for Sinhala in medical contexts.
- There are no integrated response recommendations to reduce typing effort for DHH users.

To validate these findings, interviews were performed with Deaf people to learn about their communication challenges in hospitals, as well as with sign language interpreters to uncover common misunderstandings.

Figure 4.2: Conducting User Studies

The findings validated the critical need for a mobile communication tool that can assist Sinhala speech recognition, text generation, and speech synthesis in healthcare situations.

4.2 Defining Solution Objectives

A research gap was identified, focusing on the lack of accessible, real-time communication systems that enable seamless interaction between healthcare professionals and DHH patients in Sinhala. Based on the identified gaps, the following key objectives were established. Create a mobile app that allows doctors to speak in Sinhala with real-time transcription, DHH patients to select/edit pre-generated responses (3 alternatives per query), doctors to convert patient responses to text-to-speech, and patients to enter customized queries via typing mode.

4.3 Design and Development

During this phase, we iteratively designed and refined the prototype in response to user feedback and co-design principles with Deaf and Hard of Hearing (DHH) individuals . To ensure accessibility and efficacy, the approach included several design changes, usability testing, and revisions.

4.3.1 Initial Prototype and User Feedback

In this step, a prototype of the mobile application was developed, incorporating the identified features and requirements. The prototype was designed with a focus on user-centered design principles, ensuring that the interface was simple, clear, and accessible. Early versions of the app were shared with DHH individuals for their feedback, helping to identify usability issues and gather suggestions for improvements.

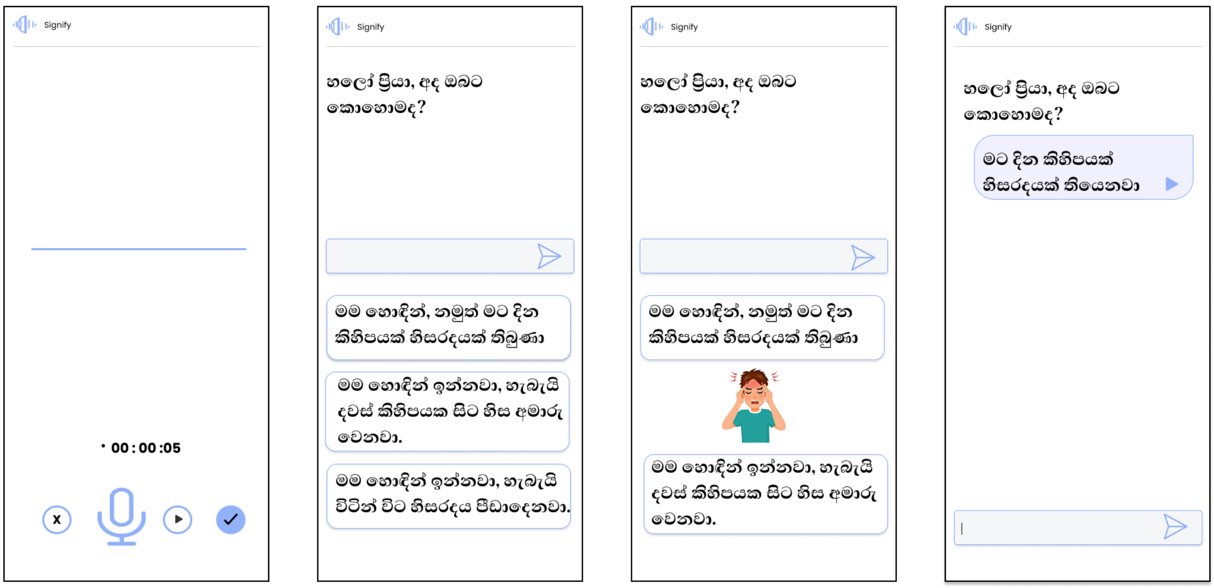


Figure 4.3: The Initial Prototypes

4.3.2 Design Refinements Based on User Preferences

Several key design decisions were made based on participant feedback:

Color Scheme Selection

The color scheme employed in this study contains several shades of blue, specifically #91B1F8, #214CAA, #081A41, #020D24, #244386, and #0E44B9, with the theme color #254B9D chosen based on feedback from DHH individuals . According to research, color selection can have a considerable impact on readability, accessibility, and user comfort, especially for users with different visual sensitivities [43]. In this scenario, #254B9D was chosen as the primary theme color because it created a balanced and visually acceptable contrast to the background colors, so improving the entire visual experience. Studies have shown that medium to dark blue tones, such as #254B9D, frequently provide ideal contrast and are simpler to notice for a wide spectrum of users, including those with visual impairments.

Font Selection

We tested the following font families. Noto Sans Sinhala, Abhaya Libre, Maname, Gemunu Libre, Yaldevi, and Stick No Bills. And based on participant suggestions, we chose "Noto Sans Sinhala." Noto Sans was chosen for this study because its design stresses legibility and readability across multiple languages, including Sinhala [43]. Unlike other typefaces, such as Stick No Bills or Gamunu, which may have more stylistic or regional focuses, Noto Sans provides a comprehensive solution for multilingual applications, providing consistency and accessibility in digital settings. According to research, fonts like Noto Sans boost user comprehension and interface usability due to their clean, sans-serif style, making them suitable for a variety of applications.

Text Display and Alignment

To make it easier for Deaf and Hard of Hearing (DHH) individuals to read, bold font weight was chosen. It has been demonstrated that bold fonts improve legibility by increasing the contrast between the text and backdrop, which makes the text easier to read and more distinct. Bold typefaces ease cognitive strain and promote faster and more accurate reading for DHH individuals, who may rely more on visual clues. Additionally, studies show that strong fonts work especially well in digital settings where readability may be impacted by screen glare and changing lighting. The content shines out more when bold text is used, which enhances accessibility and comprehension for this particular user group.

The response text was oriented left instead of centered or justified. Left-aligned text is often regarded as the most readable and easiest to digest by users, particularly in languages that read from left to right. According to studies, left alignment provides a predictable starting place for each line of text, increasing reading efficiency. While justified text might be visually appealing, it can cause irregular spacing between words, resulting in 'rivers' of white space that can disrupt reading flow, particularly for people who have difficulty reading [44]. Furthermore, centered text can result in inconsistent line lengths, making it difficult to read from line to line. As a result, left-aligned text was chosen to provide a consistent and easy-to-read style for DHH pupils, improving comprehension and accessibility.

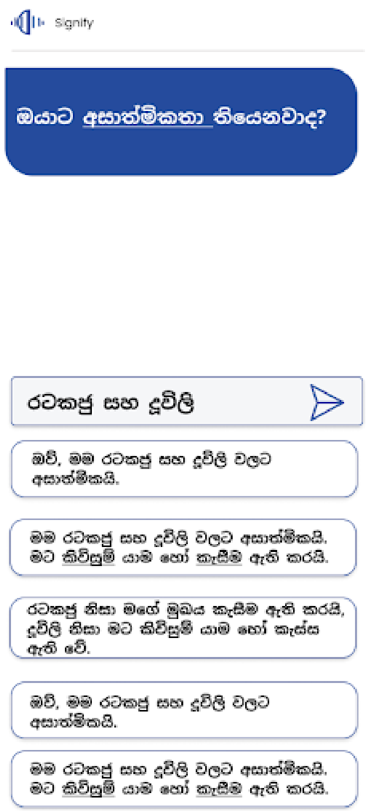


Figure 4.4: Font Weight Variations

Response Display and Layout

The decision to display three answers to a specific question in the user interface was intended to establish a balance between clarity, usability, and visual attractiveness [45]. Research has demonstrated that layout selections have a substantial impact on how users perceive and interact with interfaces. Presenting more than three solutions might result in a crowded interface, making it difficult for users to focus on individual options and creating a sensation of visual overload. This method has been demonstrated to improve cognitive processing, task completion rates, and overall interface satisfaction.

Our design uses a white background and black font for the answer and a blue background and white font for the question. Since blue backgrounds with white writing are known to increase focus and legibility, especially in low light, this method improves readability and clarity. A standard chat interface structure is followed by the question's left alignment and the answer's right alignment, which makes it easier for users to understand and less taxing on their cognitive abilities. Additionally, this design makes sure that the inquiry and response are easily distinguished, which facilitates rapid interaction and comprehension. Better visual hierarchy and simpler navigation are also ensured by utilizing contrasting colors for questions and answers.

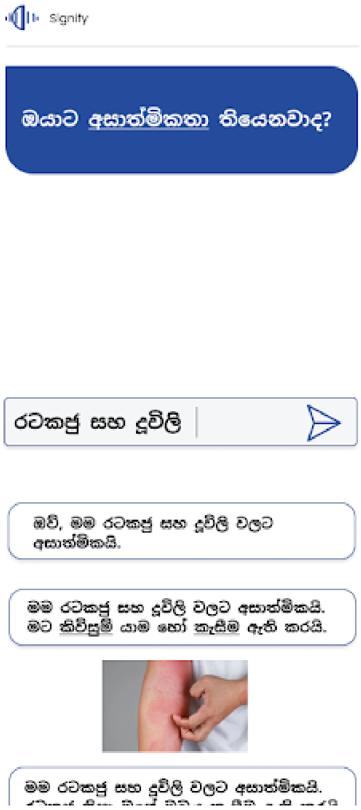


Figure 4.5: Number of Responses Displayed

4.3.3 Finalized Interface Design

After the initial prototype was developed, it was tested by DHH users in real-world scenarios, and feedback was collected through usability testing sessions. The users' suggestions were used to refine the application iteratively. This step focused on improving the user experience, resolving any usability challenges, and ensuring that the app met the expectations and needs of the DHH community. Multiple iterations were performed to continuously improve the design based on feedback.

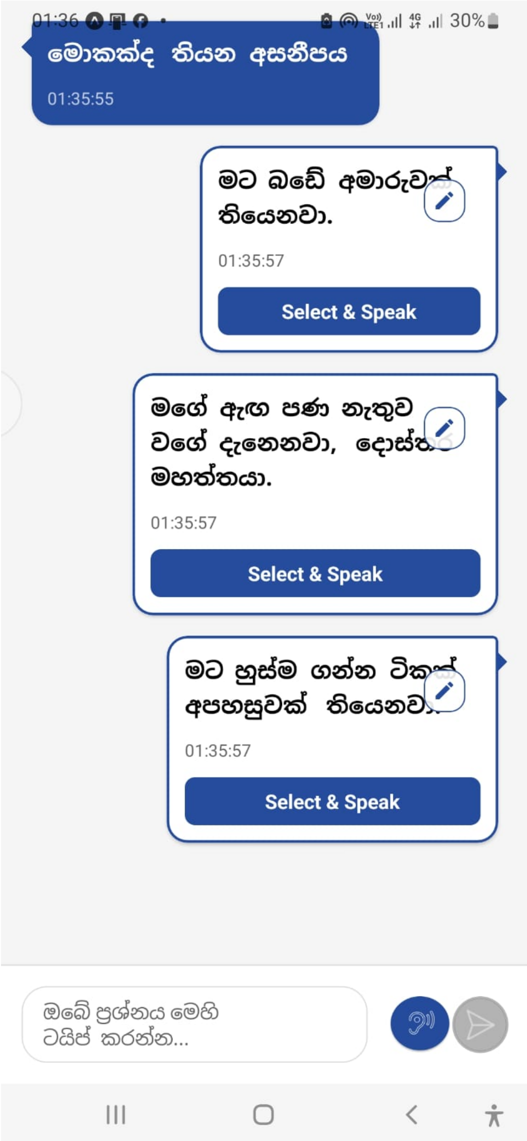


Figure 4.6: Interface if a question is asked by the doctor and responses are generated

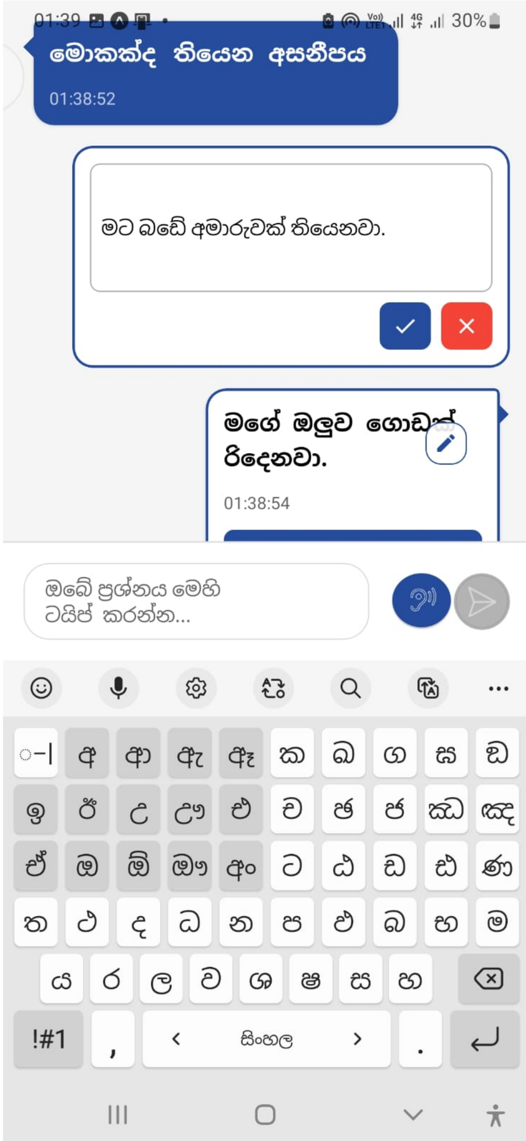


Figure 4.7: Interface of editing a response to make the reply context relevant

Figure 4.6 shows the mobile application interface when a doctor asks a question. Upon receiving the question through real-time speech-to-text conversion, the system automatically generates three potential patient responses based on the context of the question. Each generated response is displayed inside a speech bubble with an associated ”Select & Speak” button, allowing the patient to choose the most appropriate response. This design supports efficient communication by offering ready-to-use replies for the Deaf or Hard-of-Hearing (DHH) user.

Figure 4.7 illustrates the editing functionality provided by the application. If the patient wishes to modify a generated response to better match the intended meaning or context, they can tap the edit icon next to the response. This opens a text input field where the patient can edit the content. Once the modification is complete, they can confirm the edit by pressing the checkmark button. This feature ensures that communication remains accurate and personalized, giving the patient full control over the responses before they are spoken aloud.

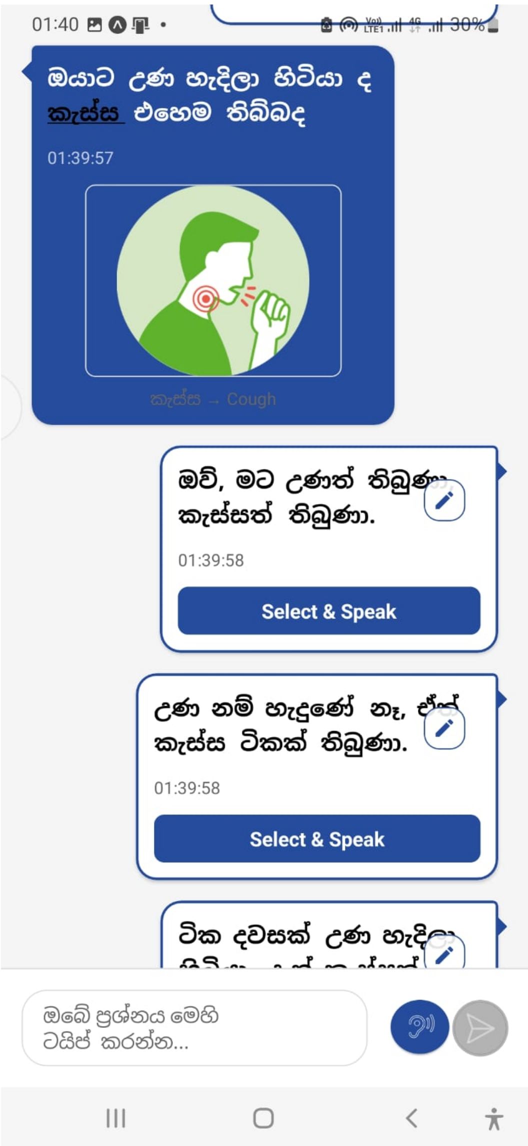


Figure 4.8: Interface of Displaying images to difficult words to understand

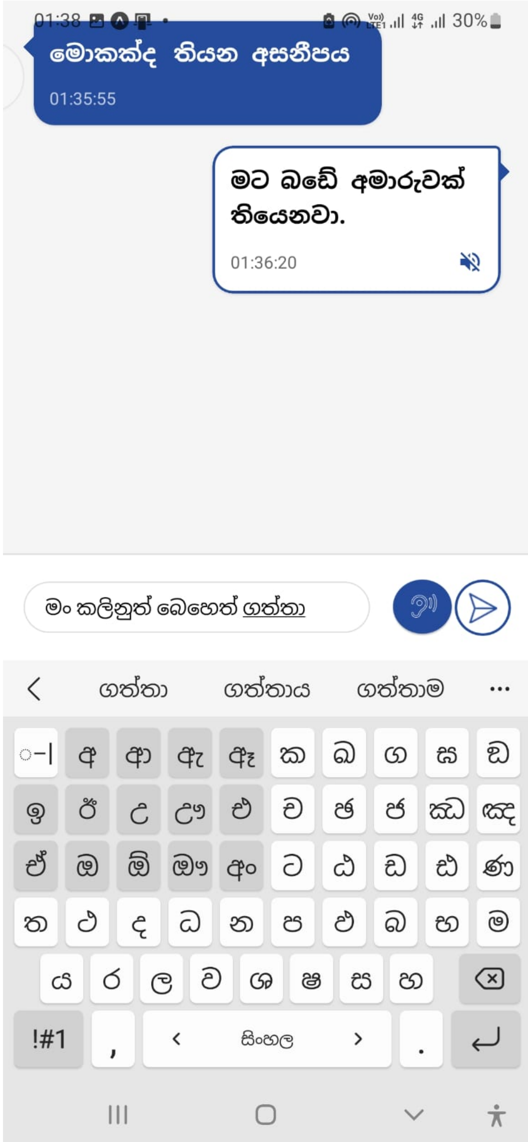


Figure 4.9: Final interface if a question is asked by the doctor and a reply is given by the deaf person

Figure 4.8 demonstrates the interface feature designed to assist users in understanding difficult medical terms. When a complex word is detected in the doctor's question or a generated response, an illustrative image along with a text label is displayed to aid comprehension. This feature is particularly valuable for Deaf and Hard-of-Hearing (DHH) users with varying literacy levels, ensuring that important concepts such as symptoms or conditions are clearly communicated through visual aids.

Figure 4.9 shows the final interaction interface where the Deaf user has either selected or typed a reply to the doctor's question. After editing or typing their own response, the patient can send the message back to the doctor. This screen emphasizes the system's support for both selecting AI-generated responses and providing fully custom responses, allowing flexible, accessible communication tailored to the user's needs.

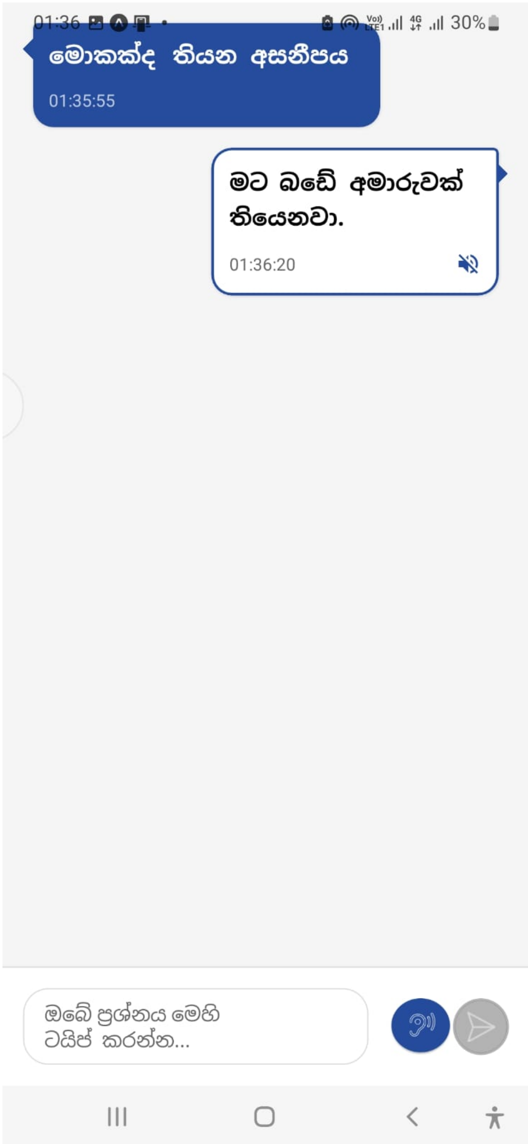


Figure 4.10: Interface of the deaf person talking back to the doctor

Figure 4.10 illustrates the final user interface of the mobile application. The interaction demonstrates the seamless communication between the doctor and deaf patient facilitated by the app, with the patient's reply appearing prominently on the screen after selection.

4.3.4 Development

The proposed mobile application was developed to facilitate seamless communication between deaf individuals and medical professionals. The system employs a combination of speech-to-text conversion, text-to-speech synthesis, and AI-driven response generation to bridge the communication gap. The implementation consists of several key components, each leveraging different APIs and machine learning models to ensure functionality and usability.

System Architecture

The application follows a client-server architecture, where the mobile frontend interacts with multiple cloud-based APIs for speech processing, text generation, and image retrieval. The primary modules include :

- Speech-to-Text Conversion
- Text Processing and Image Suggestion
- AI-Generated Response Suggestions
- Text-to-Speech Conversion

1. **Speech-to-Text Conversion**
   When a medical professional speaks, the application captures the audio input and converts it into text using Google's Speech-to-Text API. This API was selected due to its high accuracy in real-time transcription and support for multiple languages. The transcribed text is then displayed on the application interface, allowing the deaf user to read and comprehend the medical professional's statements.

2. **Text Processing and Image Suggestion**
   To enhance understanding, the application identifies complex medical terminologies or difficult words in the transcribed text. Using a combination of Google Translator API and Wikimedia, the system retrieves relevant images corresponding to these terms. This feature aids deaf users in visualizing unfamiliar concepts, ensuring better comprehension.

3. **AI-Generated Response Suggestions**
   To facilitate quick and efficient replies, the application integrates Gemini API to generate three contextually appropriate responses based on the medical professional's input. The deaf user can:

   - Select a pre-generated response as-is,
   - Edit the response for personalization, or
   - Type a completely new response.

   This flexibility ensures that communication remains natural and user-centric.

4. **Text-to-Speech Conversion**
   The user's selected or typed response is then converted into speech using Google's Text-to-Speech (TTS) API. Since Sinhala language support is unavailable in Google TTS, Malay (MY-Standard) voice was selected as the closest alternative due to its phonetic similarity to Sinhala. Although Divehi (Dhivehi) is linguistically closer, its lack of TTS support necessitated the use of Malay. The synthesized speech is played aloud, allowing the medical professional to hear the deaf individual's response.

Technical Stack

The frontend of the application is built using React-Native, ensuring cross-platform compatibility on both iOS and Android smartphones. Flask powers the backend by managing server-side logic and API queries. The system incorporates a number of APIs for essential features, including Google Text-to-Speech for natural-sounding voice synthesis, the Gemini API for intelligent response generation, and Google Speech-to-Text for precise voice transcription. Additionally, based on user input, suitable image suggestions are provided via the Google Translator and Wikimedia APIs.

4.4 Demonstration

In this phase, the mobile application was presented to users and stakeholders to showcase its functionality in real-world medical contexts. Feedback was gathered, and performance was assessed.

- App Presentation to Users and Stakeholders:
  The application was demonstrated to the Deaf/Hard-of-Hearing (DHH) individuals who provided feedback during the earlier stages of development. The demonstration focused on showing how the app transcribes speech, generates responses, allows the user to select and edit a response, and converts the final response into speech for the patient.

- User Interaction and Feedback Collection:
  During the demonstration, users interacted with the app, and their feedback was actively collected through direct observation, interviews, and questionnaires. This feedback helped identify any remaining usability issues and provided insights into areas for improvement.

- Performance and Usability Assessment:
  The demonstration also included testing the app's real-time speech-to-text transcription and the accuracy of the responses generated. Performance metrics, such as response time and task completion rate, were also assessed to ensure the app worked seamlessly in a medical context.

4.5 Evaluation

A convergent parallel mixed-method approach was employed for the research evaluation. Data collection was conducted using a cross-sectional method. Both experimental and observational designs were integrated to assess the app's effectiveness. The experiment will concern two groups: DHH individuals and hearing individuals. The comparison of the feedback by both groups will evaluate how clear, user-friendly, and communicative the app makes their experiences.

Ethical Considerations

Ethical guidelines will be strictly followed throughout the study. All participants will be explained the purpose of the study, and informed consent will be obtained with the understanding that their participation is voluntary. More than this, special measures will be taken to ensure the comfort and respect of the DHH participants by protecting their personal data.
Following implementation, the model is evaluated to ensure its effectiveness and usability. Feedback is collected from deaf users and their interactions with hearing individuals to assess the system's performance. Modifications are made based on the feedback to improve the user experience.

4.6 Communication

This phase focused on sharing the results of the project with the research community, stakeholders, and the public. It involved dissemination of the app's findings, presentations, and exploration of future collaboration and improvements.

- Results Dissemination:
  The results of the usability tests, user feedback, and overall evaluation were shared with the research community through papers and reports. These materials communicated the effectiveness of the app in addressing the communication barriers faced by DHH individuals in medical settings.

- Future Plans and Collaboration:
  The phase also involved discussing future steps for the application, such as potential enhancements, wider deployment, or further research. This includes communicating the app's scalability and its adaptability to different languages or healthcare settings.

Chapter 5

Results and Evaluation

The evaluation framework will employ a mixed-methods strategy, integrating both qualitative and quantitative assessments to evaluate user experience and the effectiveness of communication. Textual feedback will be examined to derive insights regarding the application's influence on interactions between doctors and patients, as well as to identify communication obstacles.

5.1 Performance Evaluation

Using common speech processing criteria, the effectiveness of text-to-speech (TTS) and speech-to-text (STT) systems was assessed. In order to verify the reliability of a system, these metrics evaluate accuracy, mistake rates, fluency, and naturalness.

5.1.1 Speech-to-Text Conversion

Evaluation Metrics for Speech-to-Text Conversion

To objectively evaluate the performance of the speech-to-text component, several standard evaluation metrics were used. These metrics help quantify the accuracy and quality of the transcribed output when compared to the original reference sentences. The following definitions outline each metric used in the analysis:

- **Accuracy**
  Accuracy measures the proportion of correctly recognized words over the total number of words in the reference. It is calculated using the formula:

  $$\text{Accuracy} = \frac{N - S - D - I}{N} \tag{5.1}$$

  where:

  - N is the total number of words in the reference,
  - S is the number of substitutions,
  - D is the number of deletions, and
  - I is the number of insertions.

- **Word Error Rate (WER)**
  WER is a metric that is used to evaluate automatic speech recognition systems. It represents the percentage of incorrect words and is computed as:
  $$\text{WER} = \frac{S + D + I}{N} \tag{5.2}$$
- **Character Error Rate (CER)**
  Similar to WER, but computed at the character level, CER is useful for languages where characters carry more meaning (e.g., Sinhala):
  $$\text{CER} = \frac{S + D + I}{N} \tag{5.3}$$
  where N represents the total number of characters in the reference.
- **Sentence Error Rate (SER)**
  SER considers whether an entire sentence has been transcribed correctly. A sentence is marked incorrect if there is any deviation from the reference.
  $$\text{SER} = \frac{\text{Number of Incorrect Sentences}}{\text{Total Number of Sentences}} \tag{5.4}$$
- **Match Error Rate (MER)**
  MER indicates the ratio of errors (substitutions, deletions, insertions) to the total number of words that should have been matched. It is calculated as:
  $$\text{MER} = \frac{S + D + I}{C + S + D} \tag{5.5}$$
  where CC is the number of correctly recognized words.

These metrics were applied to evaluate transcriptions of multiple Sinhala sentences under slightly noisy environmental conditions, as presented in Table 5.1: Speech-to-Text Evaluation Metrics for Sinhala Sentences.

Speech-to-Text Evaluation
To assess the accuracy of the speech-to-text component of the system, both objective and comparative evaluations were conducted using standard metrics including Accuracy, Character Error Rate (CER), Word Error Rate (WER), Match Error Rate (MER), and Sentence Error Rate (SER). A set of Sinhala reference sentences was tested across multiple attempts to ensure reliability and consistency in the results. Notably, the evaluations were performed in a moderately noisy environment with ambient sounds such as air conditioning and other background noise, but not in a human-interactive or highly chaotic setting.

As shown in Table 5.1 (located below this section), the system demonstrated excellent performance, achieving 100% Accuracy, and 0% CER, WER, SER, and MER for most test cases. This indicates that the generated transcripts were perfectly aligned with the reference sentences, even in the presence of mild environmental noise.

However, for the sentence "කොච්චර කාලයක ඉඳන්ද" , a consistent variation was observed in the transcription as "කොච්චර කාලයක ඉඳන් ද". Although this is a minor difference in phrasing, it impacted the metric scores, resulting in a drop in performance for that instance: Accuracy was reduced to 33.33%, while WER, SER, and MER increased to 66.67%, and CER to 5.56%. This highlights the system's sensitivity to subtle linguistic variations.

In summary, the results in Table 5.1 demonstrate that the speech-to-text model is robust and reliable in moderately noisy environments, and performs well with Sinhala speech in typical usage scenarios.

Speech-to-Text	**First Attempt**	**Second Attempt**	**Third Attempt**	**AVG**
Reference: මොකක්ද තියෙන අසනීපය	**"Accuracy"**: 100.0% **"CER"**: 0.0% **"MER"**: 0.0% **"SER"**: 0.0% **"WER"**: 0.0% **"Transcript"**: ""මොකක්ද තියෙන අසනීපය"	**"Accuracy"**: 100.0% **"CER"**: 0.0% **"MER"**: 0.0% **"SER"**: 0.0% **"WER"**: 0.0% **"Transcript"**: "මොකක්ද තියෙන අසනීපය"	**"Accuracy"**: 100.0% **"CER"**: 0.0% **"MER"**: 0.0% **"SER"**: 0.0% **"WER"**: 0.0% **"Transcript"**: "මොකක්ද තියෙන අසනීපය"	**"Accuracy"**: 100.0% **"CER"**: 0.0% **"MER"**: 0.0% **"SER"**: 0.0% **"WER"**: 0.0%
Reference: කලින් මොනාහරි බෙහෙත් අරන් තියෙනවද	**"Accuracy"**: 100.0% **"CER"**: 0.0% **"MER"**: 0.0% **"SER"**: 0.0% **"WER"**: 0.0% **"Transcript"**: "කලින් මොනාහරි බෙහෙත් අරන් තියෙනවද"	**"Accuracy"**: 100.0% **"CER"**: 0.0% **"MER"**: 0.0% **"SER"**: 0.0% **"WER"**: 0.0% **"Transcript"**: "කලින් මොනාහරි බෙහෙත් අරන් තියෙනවද"	**"Accuracy"**: 100.0% **"CER"**: 0.0% **"MER"**: 0.0% **"SER"**: 0.0% **"WER"**: 0.0% **"Transcript"**: "කලින් මොනාහරි බෙහෙත් අරන් තියෙනවද"	**"Accuracy"**: 100.0% **"CER"**: 0.0% **"MER"**: 0.0% **"SER"**: 0.0% **"WER"**: 0.0%

Reference: පවුලේ අයටත් තියෙනවද මේ ලෙඩේ	**”Accuracy”**: 100.0% **”CER”**: 0.0% **”MER”**: 0.0% **”SER”**: 0.0% **”WER”**: 0.0% **”Transcript”**: "පවුලේ අයටත් තියෙනවද මේ ලෙඩේ"	**”Accuracy”**: 100.0% **”CER”**: 0.0% **”MER”**: 0.0% **”SER”**: 0.0% **”WER”**: 0.0% **”Transcript”**: "පවුලේ අයටත් තියෙනවද මේ ලෙඩේ"	**”Accuracy”**: 100.0% **”CER”**: 0.0% **”MER”**: 0.0% **”SER”**: 0.0% **”WER”**: 0.0% **”Transcript”**: "පවුලේ අයටත් තියෙනවද මේ ලෙඩේ"	**”Accuracy”**: 100.0% **”CER”**: 0.0% **”MER”**: 0.0% **”SER”**: 0.0% **”WER”**: 0.0%
Reference: කොච්චර කාලයක ඉඳන්ද	**”Accuracy”**: 33.33% **”CER”**: 5.56% **”MER”**: 50.0% **”SER”**: 66.67% **”WER”**: 66.67% **”Transcript”**: කො"ච්චර කාලයක ඉඳන්ද"	**”Accuracy”**: 33.33% **”CER”**: 5.56% **”MER”**: 50.0% **”SER”**: 66.67% **”WER”**: 66.67% **”Transcript”**: කො"ච්චර කාලයක ඉඳන්ද"	**”Accuracy”**: 33.33% **”CER”**: 5.56% **”MER”**: 50.0% **”SER”**: 66.67% **”WER”**: 66.67% **”Transcript”**: කො"ච්චර කාලයක ඉඳන්ද"	**”Accuracy”**: 33.33% **”CER”**: 5.56% **”MER”**: 50.0% **”SER”**: 66.67% **”WER”**: 66.67%

Table 5.1: Speech-to-Text Evaluation Metrics for Sinhala Sentences

5.1.2 Text-to-Speech Conversion

Evaluation of Text-to-Speech (TTS) System

The evaluation of the Text-to-Speech (TTS) system was conducted primarily through a subjective listening test, designed to measure the naturalness and intelligibility of synthesized speech. The assessment was carried out in a simulated medical scenario involving doctor-patient conversations, aiming to reflect real-world application for deaf individuals who rely on TTS for communication support.

Participants

The evaluation panel included teachers from a deaf school, colleagues, and friends familiar with the needs of deaf users. This diverse group was selected to ensure a well-rounded and realistic appraisal of the TTS output in a controlled, yet practical environment.

Methodology
Listeners were asked to rate synthesized utterances on a Mean Opinion Score (MOS) scale from 1 (poor) to 5 (excellent), considering factors such as clarity, naturalness, and ease of understanding. The utterances were derived from the scripted doctor-patient dialogue scenario (see Appendix), which was designed to simulate typical communication exchanges in a medical setting.

$$\text{MOS} = \frac{\sum_{i=1}^{N} r_i}{N} \tag{5.6}$$

Results
The overall Mean Opinion Score (MOS) for the TTS system was 3.50, indicating a moderate level of performance with clear areas for improvement. When analyzed per listener group, Listener 1—primarily teachers from the deaf school—rated the system the highest with an average MOS of 4.17, reflecting a positive reception among experienced educators of the deaf community. Listener 2 and Listener 6, composed of colleagues and friends, both assigned an average MOS of 3.50, suggesting a balanced but moderately critical perspective. Listener 5 reported a slightly higher MOS of 3.67, indicating moderate satisfaction. Listener 3 and Listener 4, however, rated the system lower, with MOS values of 3.33 and 2.83 respectively, pointing to notable challenges in perception.

Discussion
These results suggest that the TTS system is capable of producing generally understandable speech, especially appreciated by users with backgrounds in deaf education. However, the variation in listener ratings highlights a need for further refinement, particularly in terms of prosody, pronunciation consistency, and clarity. Lower ratings from certain participants indicate that while the basic speech output is functional, enhancing articulation and naturalness could substantially improve the overall user experience. Continued tuning and targeted improvements could help bridge the performance gap across diverse listener groups.

5.1.3 Response Generation

The evaluation of response generation was conducted using a set of six sample questions typically asked by a doctor during a medical consultation with a deaf patient. Responses to these questions were generated using the Gemini API.
The doctor's questions used for testing were:

1. මොකක්ද තියන අසනීපය

2. කොච්චර කාලයක ඉදලා ද

3. වෙන මොනවද තියෙන අසනීප

4. වෙන බෙහෙත් මොනවහරි ගන්නවද

5. ඔයාට මොනාහරි ආසාත්මිකතාවයක් තියෙනවද

6. ඔයාට මේ ලෙඩේ කලින් හැදිල තියෙනවද

The evaluation metrics for the response generation process were calculated as follows:

1. **Average Response Time (ART):**
   This represents the mean time taken by the model to generate a single response. It is calculated by:

$$\text{ART} = \frac{1}{N}\sum_{i=1}^{N} t_i \tag{5.7}$$

   where:

   - ti is the time taken to generate the ith response,
   - N is the total number of responses.

   In the conducted experiment, the average response time was observed to be 1.406 seconds.

2. **Response Speed (RS):**
   This quantifies the number of responses generated per second and is computed as:

$$\text{RS} = \frac{N}{\sum_{i=1}^{N} t_i} \tag{5.8}$$

   In this case, the response speed was calculated to be approximately 0.711 responses per second.
   These metrics are crucial indicators of the system's practical usability, especially in real-time or near-real-time applications [46].

Discussion
The Gemini API demonstrated efficient performance with an average latency of around 1.4 seconds per response and a response generation speed close to 0.7 responses per second, aligning with typical expectations for modern large language models under moderate load [47]. Considering that the queries were in Sinhala—a low-resource language for many generative models [48]—the results indicate a commendable processing capability. The response time was sufficiently low to support interactive user experiences without noticeable delays, a critical requirement for maintaining conversational flow in dialogue systems [49], [46]. This efficiency makes the system particularly suitable for conversational health inquiry systems and chatbot-based consultations, where timely and context-aware responses are essential for usability and user trust [50].

5.2 Usability Evaluation

5.2.1 Final evaluation After System Implementation

The final evaluation was conducted after the deployment of the mobile application aimed at assisting Deaf students in medical settings. A cross-sectional mixed-methods approach was adopted to assess the effectiveness of the application.

Participants
The evaluation was conducted with a total of eight participants, comprising five Deaf individuals from a specialized school for the Deaf and three individuals from the broader community who met the following inclusion criteria:

- Being diagnosed as deaf.
- Having moderate literacy skills, defined as the ability to understand and read written content.
- Providing informed consent to participate in the study.

The age, gender, and hearing ability distribution of the participants is as follows:

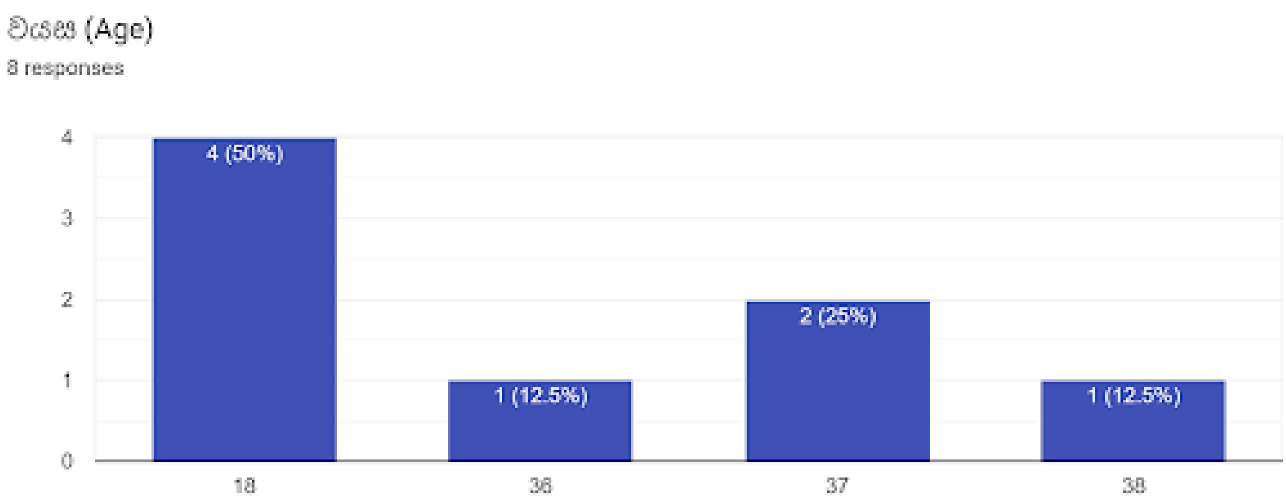


Figure 5.1: Age Distribution of Participants

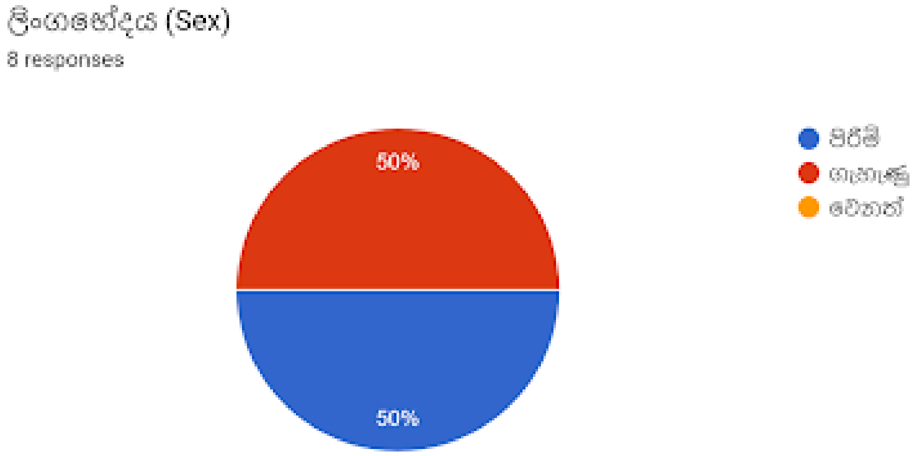


Figure 5.2: Sex Distribution of Participants

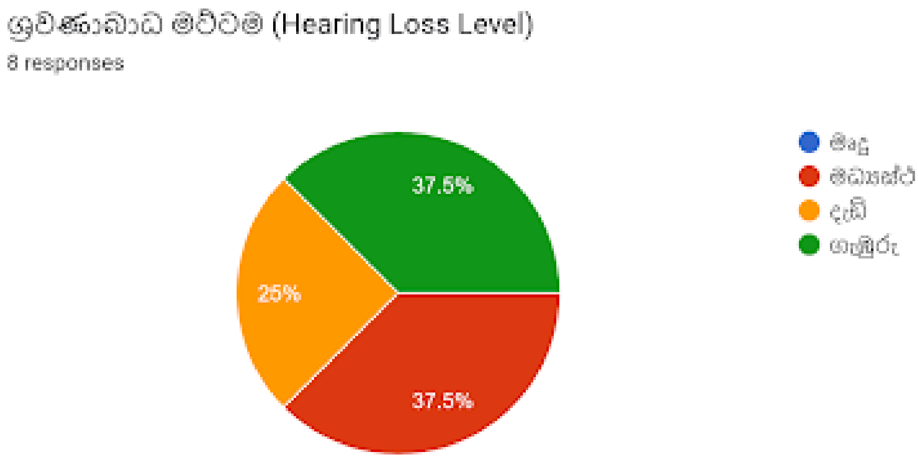


Figure 5.3: Hearing Loss Level Distribution of Participants

Data Collection Methods
The following data collection methods were employed:

- **Surveys and Questionnaires**
  Structured surveys were distributed to assess user satisfaction and perceptions of the system post-deployment. These were designed to collect feedback on various aspects of the application's usability.

- **Interviews**
  Semi-structured interviews were conducted with Deaf participants to gather qualitative insights into their experiences using the application. This method enabled participants to express their thoughts freely, allowing the research team to gain a deeper understanding of the application's strengths, limitations, and areas for improvement from the perspective of the end users.

- **System Usage Data**
  System analytics were collected to monitor how users interacted with the application. Data such as usage frequency, feature engagement, and session duration were recorded to help evaluate overall user engagement and identify common usage patterns.

- **Simulation Sessions**
  Participants engaged in simulation sessions designed to reflect real-world scenarios in which the application would be used. These sessions allowed for direct observation of user behavior and provided practical insights into the system's performance, usability, and reliability in realistic conditions.

Results
After the participants had used the system for the designated period, we conducted interviews, gathered data through questionnaires, and analyzed system usage data. This comprehensive approach allowed us to gather valuable insights from participants regarding the effectiveness of the user interface (UI) and its support for the Sinhala language.

5.2.2 Analysis of Participant Feedback on User Interface and application

The analysis of participant feedback on the user interface explores which aspects have proven most effective in facilitating user engagement and comprehension. Participants were tasked with evaluating the ease of use of the UI and the adequacy of Sinhala language support.

i. **Ease of Use**
   All eight participants reported that the user interface was easy to use, indicating that the design and navigation were intuitive and accessible for Deaf individuals. This feedback suggests that the system's UI effectively accommodates the needs of users with hearing impairments.

ii. **Sinhala Language Support**
   Regarding the Sinhala language support, the responses varied:
   All participants confirmed that the Sinhala language support was sufficient for their understanding.

   These responses highlight the importance of continuous refinement in language support to cater to the diverse linguistic needs of users.

   All participants responded affirmatively to both questions, indicating a unanimous positive perception of the system's usability and language support capabilities.

   To visually represent these findings, pie charts (**??** & **??**) illustrate the proportion of positive responses. Given the uniformity of the responses, the chart displays a single segment representing 100% affirmative feedback for both UI usability and Sinhala language

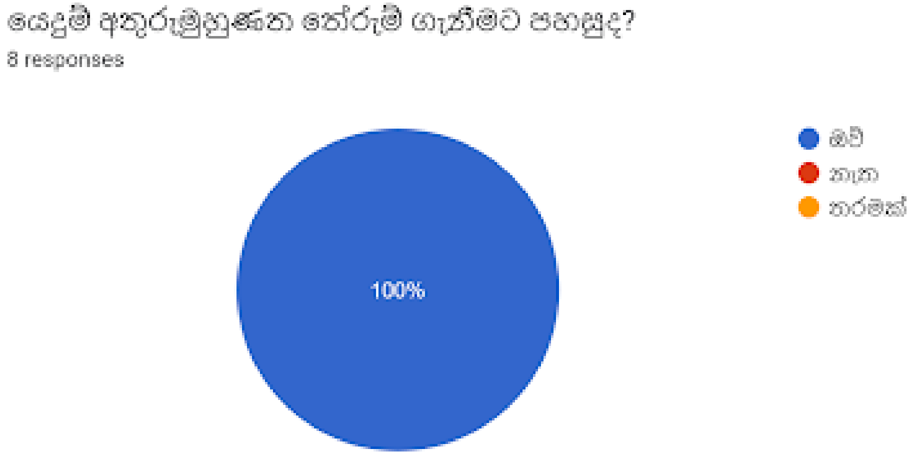


Figure 5.4: Participant Feedback on UI Usability and Sinhala Language Support

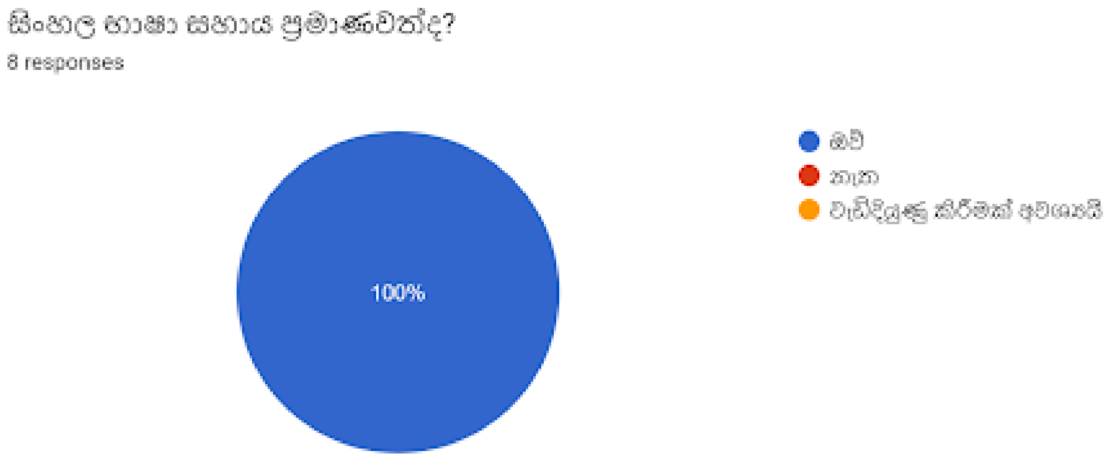


Figure 5.5: Participant Feedback on Sinhala Language Support

support.

This unanimous positive feedback underscores the system's effectiveness in delivering an accessible and user-friendly experience for Deaf users, with particular emphasis on the clarity and usability of the interface and the adequacy of Sinhala language support.

iii. **Participant Evaluation of Speech-to-Text Usability**
Participants were asked to rate the accuracy of the speech-to-text feature on a 5-point Likert scale, where 1 indicates "Very Poor" and 5 indicates "Excellent." The distribution of responses was as follows:

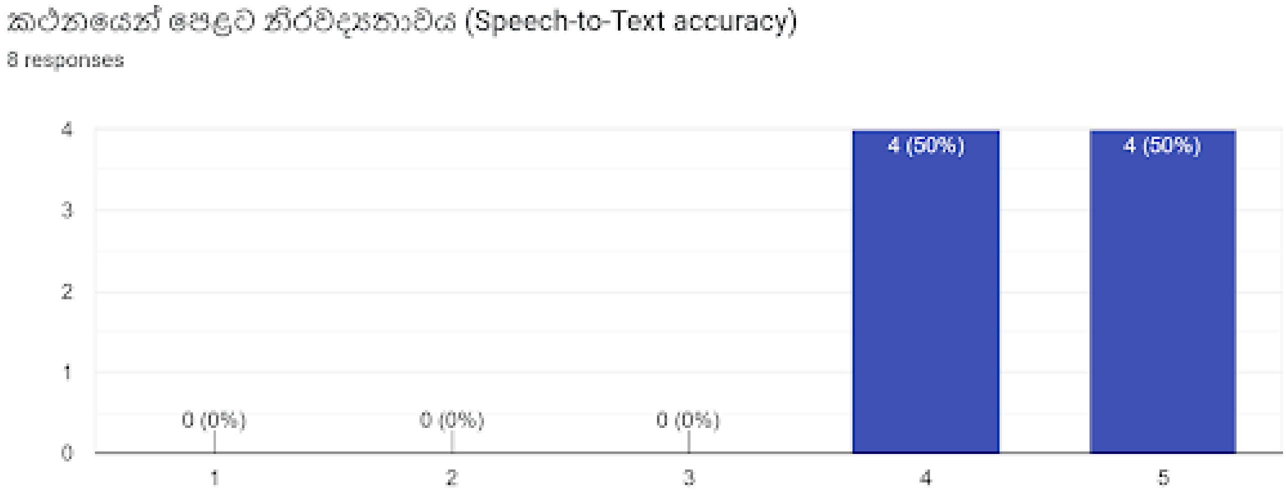


Figure 5.6: Participant Feedback on Speech-to-Text

The evaluation revealed that 50% of participants rated the STT accuracy as 5, denoting an "Excellent" performance, while the other 50% assigned a rating of 4, indicating a "Good" performance. This distribution of responses demonstrates a high level of satisfaction with the STT functionality, suggesting that the feature effectively meets the communication needs of the users. To visually represent these findings, a bar chart is employed in Figure 5.6, illustrating the proportion of each rating and providing a clear depiction of the overall positive reception of the STT feature.

iv. **User Feedback on Text-to-Speech Feature**

The results indicate that the majority of participants rated the text-to-speech (TTS) feature very favorably, with 75% assigning it the highest rating of 5, corresponding to an "Excellent" level of performance. The remaining 25% of participants gave it a rating of 4, which signifies a "Good" level of effectiveness. This pattern of responses demonstrates a generally strong approval of the TTS functionality, suggesting that it successfully delivers clear and intelligible speech output. Figure 5.7 presents a bar chart that visually summarizes these ratings, providing a clear overview of the positive user perceptions towards the TTS feature.

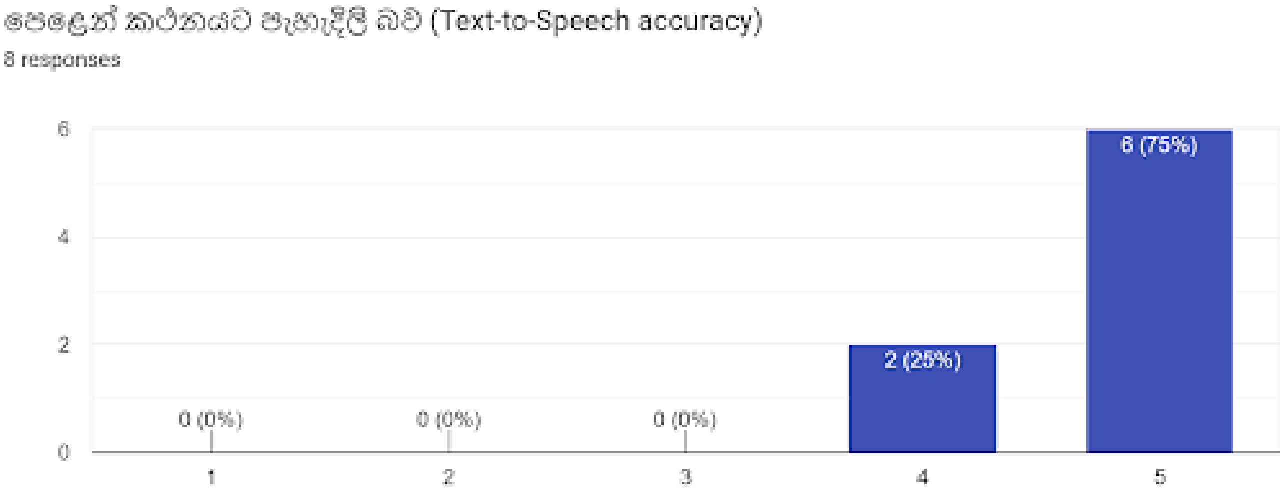


Figure 5.7: Participant Feedback on Text-to-Speech

v. **User Feedback on Response Generation Feature**
To assess the effectiveness of the response generation feature, participants were asked to rate its performance on a five-point scale. The results indicate a varied but generally positive evaluation: 25% of participants rated the feature as "Average" (3), 37.5% assigned a "Good" rating (4), and another 37.5% rated it as "Excellent" (5). This distribution suggests that while most users found the feature to perform well, a subset perceived room for improvement. Overall, the ratings reflect a satisfactory level of user acceptance and indicate that the response generation functionality largely meets user expectations in facilitating effective interactions.

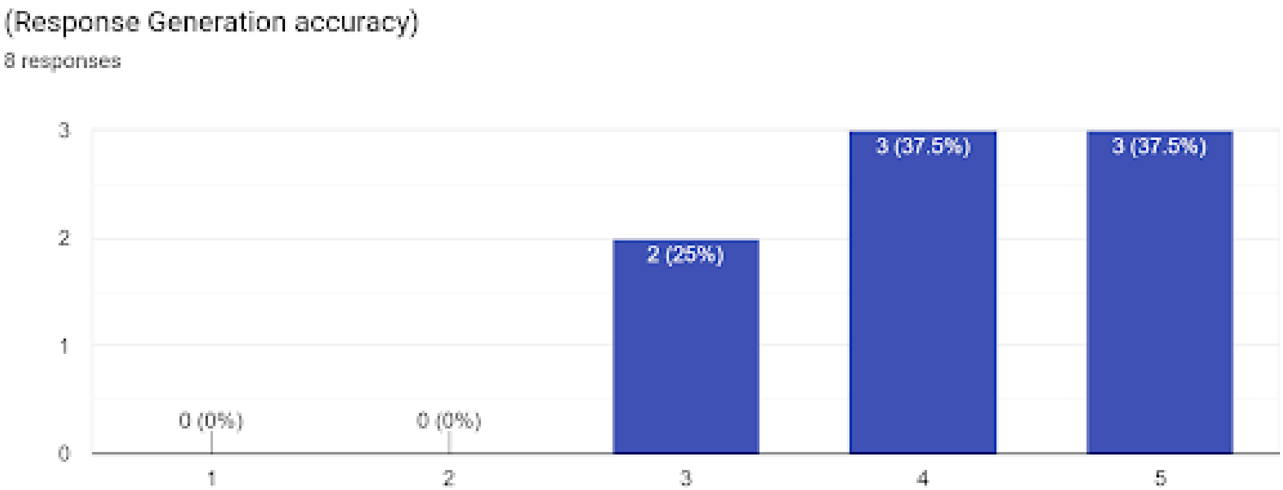


Figure 5.8: Participant Feedback on Response Generation

vi. **User Feedback on Image Generation Feature**
Participants were asked to evaluate the image generation feature, which provides visual support for difficult words. The evaluation results showed that 75% of participants rated

the feature as "Excellent" (5), while 12.5% assigned it a "Good" rating (4), and another 12.5% rated it as "Poor" (2). The participant who provided the lowest rating demonstrated a high literacy level and indicated that visual aids were less necessary for their comprehension. Despite this, the overall positive ratings suggest that the feature is effective in enhancing understanding and accessibility, particularly for users with lower literacy levels or those who benefit from visual learning strategies.

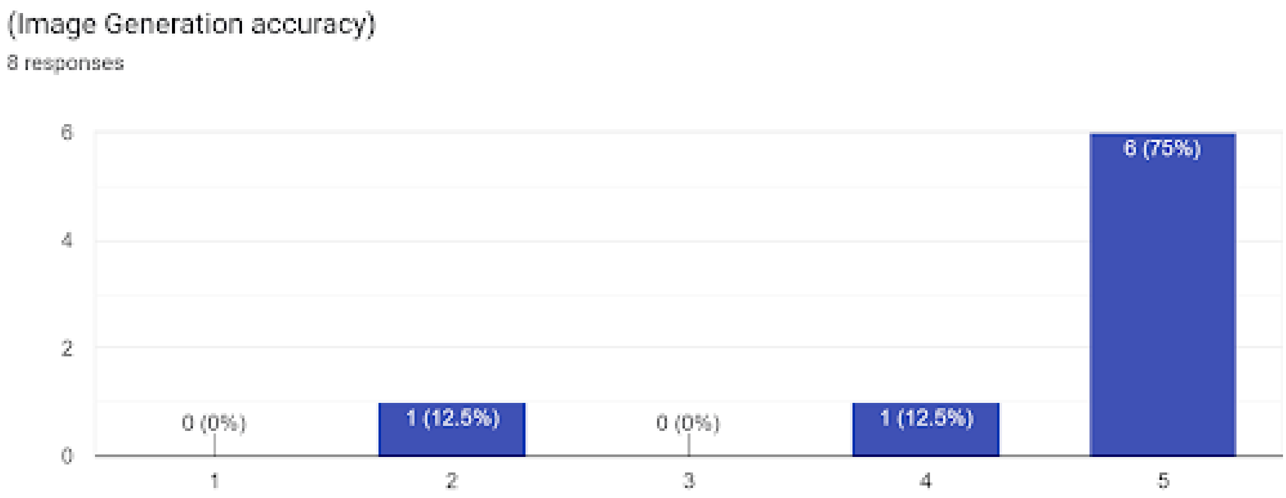


Figure 5.9: Participant Feedback on Image Generation

vii. **User Feedback on frequently used feature**

When participants were asked to identify the feature they used most frequently, half of them (50%) reported preferring the image generation feature. This preference suggests that visual aids play a significant role in supporting comprehension, particularly for users who benefit from additional graphical context. The text-to-speech feature was the next most frequently used, selected by 25% of participants, indicating its importance for auditory reinforcement. Both the speech-to-text and response generation features were favored by 12.5% of participants each, reflecting more specialized or situational usage patterns. These findings highlight the varied needs of users and emphasize the value of offering multiple complementary features to accommodate diverse communication preferences.

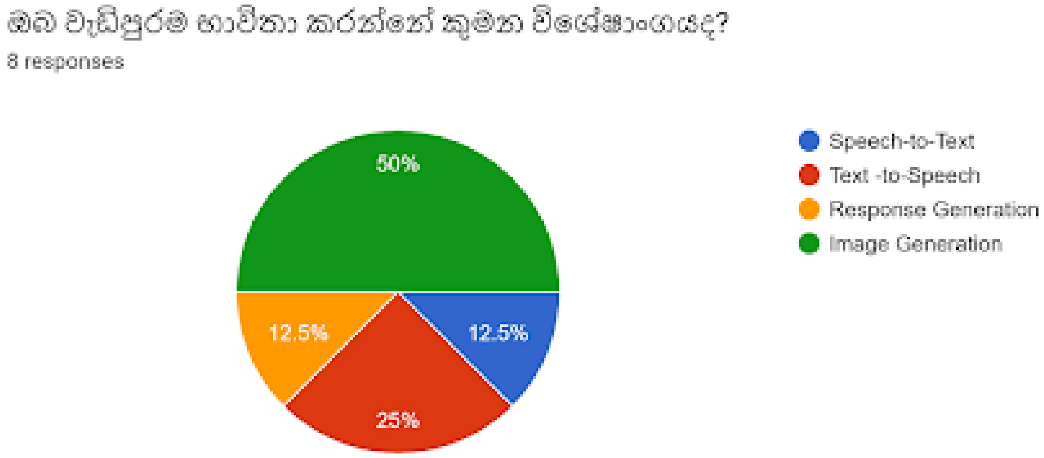


Figure 5.10: Participant Feedback on most frequently used feature

viii. **User Feedback on challenges**
The majority of participants (75%) expressed concerns related to translation accuracy, specifically noting uncertainty about whether the translated content accurately reflected the doctor's original message and whether the text they entered was correctly conveyed back to the doctor. This ambiguity affected users' confidence in the communication process.No participants reported issues with slow system responses or user interface difficulties. It is important to note that some Deaf participants naturally required additional time to process information due to their communication style. Additionally, 25% of users raised other concerns, such as the absence of sign language support and a preference for simpler, shorter sentences for the response generation feature instead of longer, more complex ones. These findings highlight key areas for improvement to increase the system's reliability and user trust.

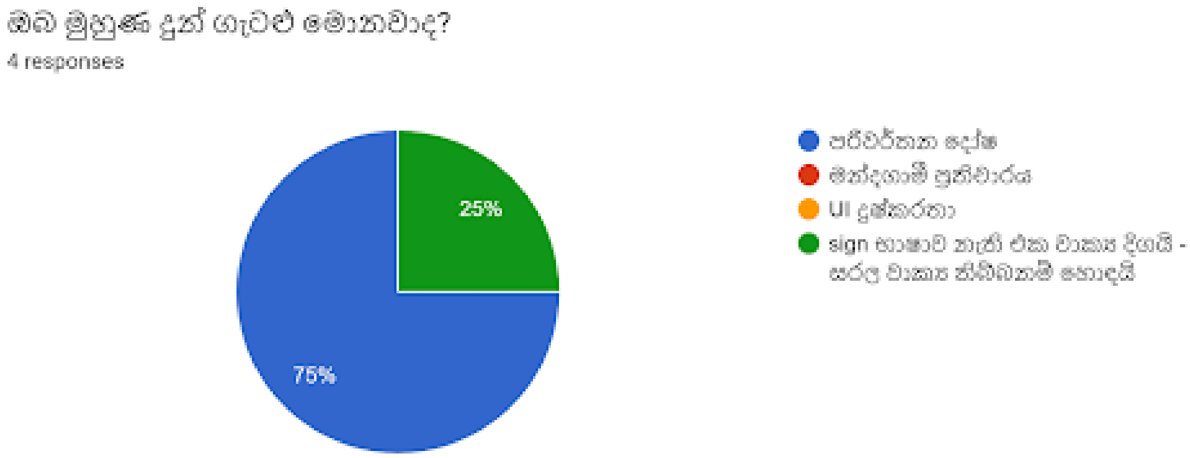


Figure 5.11: Participant Feedback on challenges

ix. **User Feedback on improving the application**
Participants provided valuable suggestions on how the system could be further improved. Forty percent of respondents emphasized the need for enhanced translation accuracy to

better capture the intended meaning during communication. Another 40% expressed a desire for expanded support for the Sinhala language, indicating that broader language coverage would increase the system's usability. None of the participants identified the need for faster processing speeds as a priority. Additionally, 20% of users recommended incorporating sign language support as a visual alternative to images, suggesting that this could further enhance accessibility and user experience. These insights offer clear directions for future development efforts to better meet the needs of the target user group.

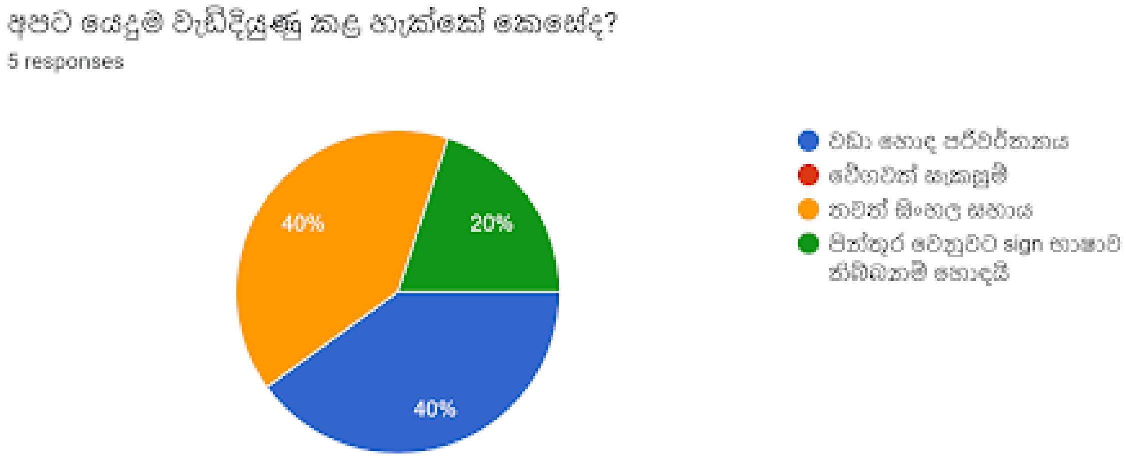


Figure 5.12: Participant Feedback on improving the system

x. **User Feedback on usage and advantages of the application**
All participants (100%) affirmed that the application is useful in their daily lives, indicating its practical value and relevance to their communication needs. Furthermore, a substantial majority of participants (87.5%) expressed their intention to continue using the application over time, reflecting sustained user acceptance and satisfaction. No participants indicated that they would discontinue use or were uncertain about continued engagement with the system.

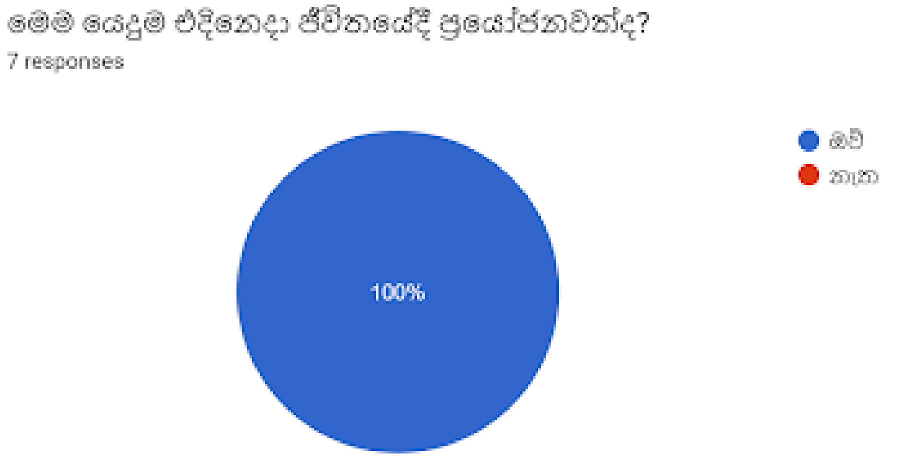


Figure 5.13: Participant Feedback on system as a daily necessity

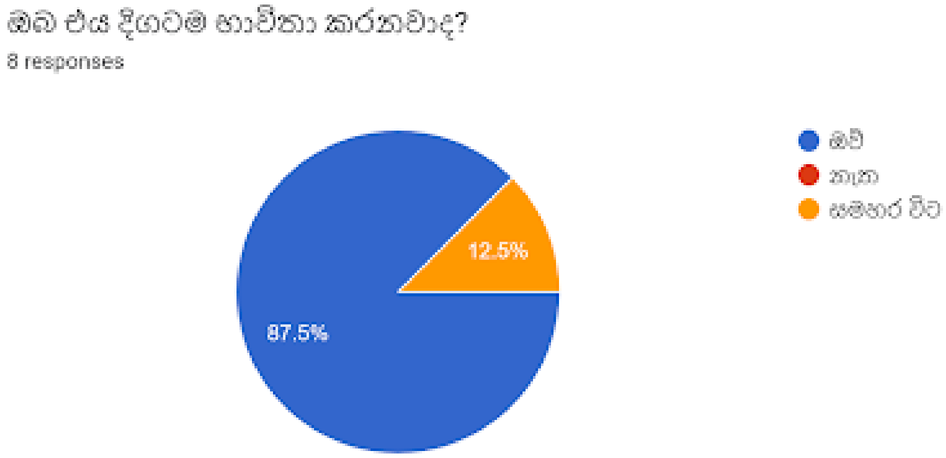


Figure 5.14: Participant Feedback on usage of the system in a daily basis

xi. **User Feedback on overall satisfaction**

The overall satisfaction levels reported by participants indicate a strong positive response to the application. A majority of 75% rated their experience with the system at the highest satisfaction level of 5, signifying excellent approval of its features and functionality. The remaining 25% provided a rating of 4, indicating that while their experience was generally positive, there may be minor areas for improvement. Collectively, these ratings reflect a high degree of user acceptance and suggest that the application effectively addresses the communication needs of the participants. This favorable reception underscores the system's potential to serve as a valuable tool in facilitating communication for Deaf users. However, the presence of less-than-perfect ratings also points to opportunities for further refinement to enhance usability and performance.

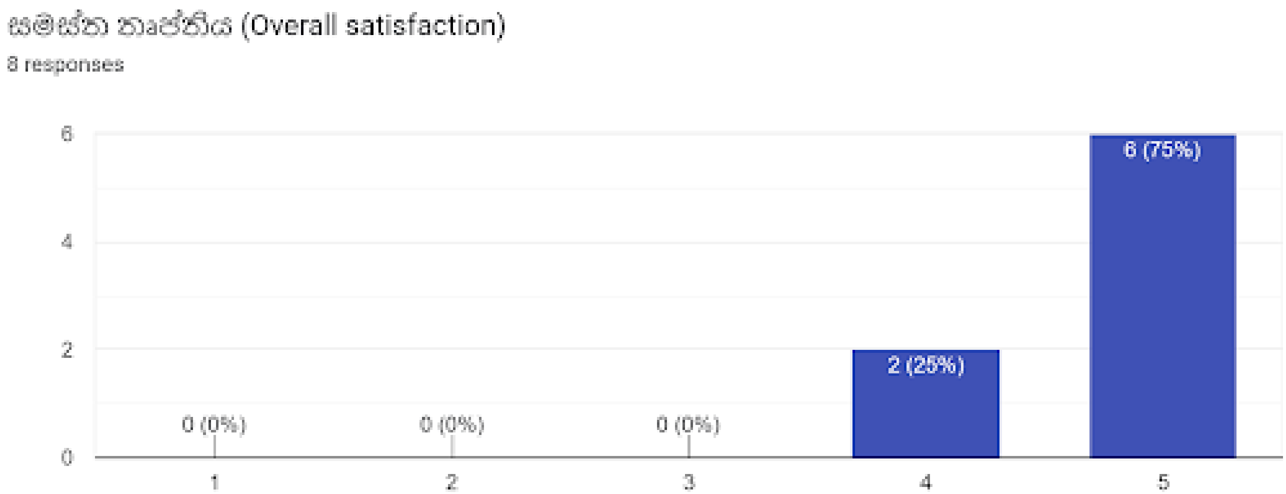


Figure 5.15: Participant Feedback on challenges

5.2.3 Evaluation of Task Completion and Efficiency

To assess user performance and the system's usability, we measured the Task Completion Rate and the time taken to complete each task during in-person usability testing. Each participant was given six tasks to complete, and for every task, we recorded both the time taken to read the instruction and the time taken to respond. The goal was to evaluate how intuitive and efficient the system was for users with hearing impairments.

All participants successfully completed all six tasks, resulting in a 100% Task Completion Rate, which reflects the high usability and accessibility of the application interface.

Task Completion Rate (%)

$$\text{Task Completion Rate} = \left(\frac{\text{Number of tasks completed successfully}}{\text{Total tasks attempted}} \right) \times 100$$

Average Time per Task

For each task:

$$\text{Total Time} = \text{Reading Time} + \text{Response Time}$$

Participant	**Tasks Completed**	**Total Tasks**	**Completion Rate (%)**	**Total Time**	**Avg Time per Task**
1	6	6	100%	66 sec	11.00 sec
2	6	6	100%	97 sec	16.17 sec
3	6	6	100%	78 sec	13.00 sec
4	6	6	100%	72 sec	12.00 sec
5	6	6	100%	65 sec	10.83 sec
6	6	6	100%	60 sec	10.00 sec
7	6	6	100%	60 sec	10.00 sec
8	6	6	100%	58 sec	9.67 sec

Table 5.2: Summary of Task Completion and Time Metrics

The average task completion time among participants ranged from 9.67 seconds to 16.17 seconds, demonstrating that users could complete tasks quickly and efficiently. Participant 2 required the most time, averaging approximately 16.17 seconds per task, while Participant 8 completed tasks the fastest with an average of 9.67 seconds. Despite these differences, all participants achieved a 100% task completion rate, reflecting the system's clarity, ease of navigation, and alignment with users' cognitive and motor capabilities.

These findings support the effectiveness of the application's interface and instruction design, confirming that it is both intuitive and accessible to users with hearing impairments.

Chapter 6

Discussion

6.1 Discussion

This research aims to enhance communication between deaf/hard-of-hearing (DHH) individuals and non-hearing impaired individuals, particularly in medical consultations in Sri Lanka. To facilitate this, we have developed and introduced a mobile application that converts speech to text, generates contextual responses, integrates visual aids and convert text to speech. Our study evaluates the application's effectiveness in addressing three key research questions.

The first research question we aimed to answer is "**How to design and develop a user-friendly communication application for deaf and hard-of-hearing individuals with hearing individuals?**"
Designing and developing a user-friendly communication application for deaf and hard-of-hearing individuals with hearing individuals required a thoughtful approach to accessibility and inclusivity. The core idea was to create an application that would bridge the communication gap in real time while being intuitive and easy to use. The app integrates Sinhala speech-to-text technology, converting spoken language into readable text for deaf users, allowing them to follow conversations without the need for lip-reading or sign language. To ensure ease of use for both deaf and hearing individuals, the user interface was designed to be simple, with clear, legible fonts and easy navigation. It also includes a text-to-speech feature, so non-hearing impaired users can hear responses in Sinhala, and also read the text for communication. Additionally, the app incorporates response generation, which suggests appropriate replies based on context. Moreover, the app integrate an image generation functionality to generate images for difficult words. The interface was kept minimalistic, with a focus on clarity and quick access to key features. The combination of these elements resulted in an inclusive, seamless experience that allows both deaf and hearing individuals to communicate effectively without barriers. .

The second research question was focused on answering, "**How to reduce communication effort for deaf or hard-of-hearing individuals?**"
Reducing the communication effort for deaf and hard-of-hearing individuals was addressed by implementing several key features. The primary method was the incorporation of speech-to-text and text-to-speech capabilities. The real-time transcription of spoken Sinhala into text allows users to follow conversations without needing to lip-read or ask for repetition. Additionally, the app generates suggested responses for common conversational scenarios, significantly reducing the effort required by the user to type out responses manually. By enabling these automated responses, the app removes the burden of constantly thinking about what to type, thus minimizing the effort. Another significant feature is the use of visual aids such as images for difficult words in the generated text, providing immediate clarification without requiring the user to search for definitions or ask for further explanation. This multimodal approach enhances

comprehension and reduces the strain of verbal communication, making it more accessible for deaf and hard-of-hearing individuals.

The third research question was focused on answering, " **How to improve the interaction speed to support real-time conversations with hearing individuals?**"
The application is designed to enhance interaction speed and support real-time conversations by minimizing delays in the communication process. When users speak, the app quickly converts the spoken Sinhala into text and displays it almost immediately, allowing the deaf or hard-of-hearing individual to follow the conversation in real time. The app also suggests responses for deaf individuals, enabling them to use appropriate replies quickly and reducing the time spent composing responses. It optimizes the response generation process for efficiency, ensuring users can communicate without unnecessary delays. Additionally, the app provides fast text-to-speech conversion, allowing hearing individuals to immediately hear the generated responses, ensuring smooth communication between users. To further enhance comprehension, the app generates images for difficult words in the responses or converted speech. These visual aids reduce the time spent searching for meanings externally, helping users understand the conversation faster. By streamlining the entire process and incorporating image generation, the app ensures a seamless conversation flow, minimizes disruptions, and supports natural, real-time interactions.

The final research question that this research study aims to answer is, "**How can the use of suggested responses and visual aids (such as images) improve the clarity of text-based communication and enhance social engagement for deaf or hard-of-hearing individuals in conversations with hearing individuals?**"
This study found that suggested responses and visual aids significantly improved communication clarity between DHH users and hearing individuals. Context-aware reply options reduced misunderstandings by allowing users to select appropriate responses quickly. Visual aids, especially images for complex terms, provided additional support, particularly benefiting users with lower literacy levels. Together, these features enhanced the accuracy and speed of text-based communication during consultations.
In addition,the application also strengthened social engagement by making conversations more fluid and less effortful for DHH users. Suggested responses helped maintain a natural dialogue flow, while visual aids made interactions more intuitive and less intimidating. Participants reported feeling more confident and comfortable using the application and expressed a strong willingness to adopt it in everyday communication scenarios.

6.2 Recommendation

Based on the study's results and feedback from users, the following suggestions are provided. These recommendations are meant to enhance the existing use of assistive communication technologies for the deaf and hard-of-hearing (DHH) community and direct future research in this area.

Deaf people with low reading levels who had trouble understanding sentences were a problem. First, the doctor's question appeared hard for deaf people to fully understand the meaning of the phrases. Therefore, in order to understand the question asked, they require the assistance of a communication facilitator. Additionally, they recommended utilizing an image that represented the word in sign language, which raised their degree of involvement even though the computer generates images for challenging words.

Furthermore, since they are deaf and must maintain concentration during the talk, the deaf people had trouble comprehending when the doctor had completed asking the question at hand. It was recommended to employ certain mechanisms, such as vibrations or automatic voice recognition, to indicate that the doctor has completed speaking.

6.3 Future Work

The current application assists Sinhala-speaking DHH individuals in medical settings, but future work will broaden its scope and capabilities. We plan to expand the system into new domains like education and retail, while adding multilingual support (e.g., Tamil, English) to serve a wider audience. A conversation history feature will allow users to revisit past interactions, and the text-to-speech (TTS) output will be enhanced with a more natural Sinhala voice. Additionally, speech recognition will be optimized for noisy environments to improve reliability in real-world settings.

Further improvements include context-aware response suggestions, customizable accessibility options (e.g., adjustable text size, gesture controls), and offline functionality for low-connectivity areas. Large-scale user testing will validate the system's adaptability across diverse scenarios, and performance optimizations will ensure compatibility with low-end devices. These advancements aim to transform the application into a comprehensive, inclusive communication tool for DHH individuals in everyday life.

Bibliography

[1] Y. Qin and X. Zhang, "HMM-based speaker emotional recognition technology for speech signal," *Adv. Mater. Res.*, vol. 230–232, pp. 261–265, 2011.

[2] A. H. Ziadat and A. A. Al Rahmneh, "The learning, social, and economic challenges facing the deaf and hearing-impaired individuals," *Cypriot Journal of Educational Sciences*, vol. 15, pp. 976–988, 2020. [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=3826721

[3] Y. Kim, J. Lee, H. Kim, and S. Lee, "Accessible healthcare for people with hearing loss: A review of current technologies and future directions," *Journal of Medical Internet Research*, vol. 26, no. 1, p. e45678, January 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949916X24000860

[4] M. Rannefeld, K. Hüning, L. Möller, and D. Jansen, "Healthcare communication barriers for deaf patients: A qualitative study on the perspectives of both patients and providers," *BMC Health Services Research*, vol. 23, no. 1, p. 256, March 2023. [Online]. Available: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-023-09154-5

[5] E. Kaplunov, "Mistrust between deaf patients and hearing staff in healthcare settings," *Empedocles: Eur. J. Philos. Commun.*, 2023.

[6] World Health Organization. (2023) Deafness and hearing loss. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[7] T. K. Holcomb, *Introduction to Deaf Culture*, ser. Professional Perspectives on Deafness Series. Oxford University Press, 2023. [Online]. Available: https://books.google.lk/books?id=r_PBEAAAQBAJ

[8] J. Koncovy, "Exploring deafhood: Investigating the experiences of d/deaf and hard of hearing people who faced barriers to deaf culture," 2023.

[9] T. Most, "The effects of degree and type of hearing loss on children's performance in class," *Deafness Educ. Int.*, vol. 6, no. 3, pp. 154–166, 2004.

[10] H. E. Obutor, "Assistive listening devices for individuals with hearing impairment," *Fac. Nat. Appl. Sci. J. Math. Sci. Educ.*, vol. 5, no. 2, pp. 78–81, 2023. [Online]. Available: https://fnasjournals.com/index.php/FNAS-JMSE/article/view/390

[11] E. M. Picou, "Hearing aid benefit and satisfaction results from the MarkeTrak 2022 survey: Importance of features and hearing care professionals," *Semin. Hear.*, vol. 43, no. 4, pp. 301–316, Nov. 2022.

[12] K. S. Rakhimova, "Cochlear implantation: An innovation in the development of technology, medicine, deaf pedagogy, and speech therapy," *Web Sci.*, vol. 4, no. 2, pp. 321–330, 2023.

[13] R. Ladner. (2010) Technology for deaf people. [Online]. Available: https://courses.cs.washington.edu/courses/cse590w/10sp/deaf-tech10.pdf

[14] K. Papadopoulos, E. Koustriava, L. Isaraj, E. Chronopoulou, F. Manganello, and R. Molina-Carmona, "Assistive technology for higher education students with disabilities: A qualitative research," *Digital*, vol. 4, no. 2, pp. 501–511, 2024.

[15] A. Connelly and J. Doyle, "Speech to text, text to speech: Support for deaf and hard of hearing students," *Odyssey: New Directions Deaf Educ.*, vol. 23, pp. 64–67, 2023. [Online]. Available: https://eric.ed.gov/?id=EJ1408651

[16] H. Kim, H. Hwang, S. Gwak, J. Yoon, and K. Park, "Improving communication and promoting social inclusion for hearing-impaired users: Usability evaluation and design recommendations for assistive mobile applications," *PLoS ONE*, vol. 19, no. 7, p. e0305726, Jul. 2024.

[17] P. Kumar, C. Venkatesan, S. A. R, D. Balamurugan, and V. Sathiyapriya. (2023) Speech-to-text transfiguration in language numerals for perpetual deaf patients. [Online]. Available: https://www.researchgate.net/publication/370715938

[18] A. Olubukola. (2021) A text to speech and speech to text application for students with hearing and speaking impairments. [Online]. Available: https://www.academia.edu/76409843

[19] L. Pragt, P. van Hengel, D. Grob, and J.-W. A. Wasmann, "Preliminary evaluation of automated speech recognition apps for the hearing impaired and deaf," *Front. Digit. Health*, vol. 4, 2022.

[20] R. Tutunov, A. Grosnit, J. Ziomek, J. Wang, and H. Bou-Ammar. (2023) Why can large language models generate correct chain-of-thoughts? [Online]. Available: https://arxiv.org/abs/2310.13571

[21] J. Wu, X. Wu, and J. Yang, "Guiding clinical reasoning with large language models via knowledge seeds," in *Proc. 33rd Int. Joint Conf. Artif. Intell. (IJCAI)*, 2024, pp. 7429–7437.

[22] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa. (2022, May) Large language models are zero-shot reasoners. [Online]. Available: https://arxiv.org/abs/2205.11916

[23] (2024) Healthcare voice AI assistants: Factors influencing trust and intention to use. [Online]. Available: https://arxiv.org/abs/2401.04543

[24] S. Huffman, S. Chen, K. A. Mack, H. Su, Q. Wang, and R. Kushalnagar. (2024, Oct.) 'We Do Use It, But Not How Hearing People Think': How the Deaf and Hard of Hearing community uses large language model tools. [Online]. Available: https://arxiv.org/abs/2410.21358

[25] S. Cheng, S. Huffman, Q. Zhu, H. Su, R. Kushalnagar, and Q. Wang. (2024, Sep.) 'Real Learner Data Matters': Exploring the design of LLM-powered question generation for Deaf and Hard of Hearing learners. [Online]. Available: https://arxiv.org/abs/2410.00194

[26] S. Frolov, T. Hinz, F. Raue, J. Hees, and A. Dengel, "Adversarial text-to-image synthesis: A review," *Neural Netw.*, vol. 144, pp. 187–209, Dec. 2021.

[27] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *Proceedings of the 38th International Conference on Machine Learning (ICML)*. PMLR, 2021, pp. 8821–8831, [Online]. Available: https://proceedings.mlr.press/v139/ramesh21a.html?ref=journey.

[28] Y. Samaradivakara and et al. (2025, Jan.) Tailored real-time AR captioning interface for enhancing learning experience of Deaf and Hard-of-Hearing (DHH) students. [Online]. Available: https://arxiv.org/abs/2501.02233

[29] Y. Perera, N. Jayalath, S. Tissera, O. Bandara, and S. Thelijjagoda, "Intelligent mobile assistant for hearing impairers to interact with the society in Sinhala language," in *Proc. 2017 IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, Dec. 2017, pp. 1–6.

[30] L. S. K. Udugama, R. Nethsinghe, J. Southcott, S. Kularathna, L. Dhanapala, and K. A. C. Alwis, "Sign language usage of deaf or hard of hearing Sri Lankans," *J. Deaf Stud. Deaf Educ.*, vol. 29, no. 2, pp. 187–198, Dec. 2023.

[31] B. G. J. Gamage, R. P. S. D. Paranagama, R. M. S. H. Ranaweera, A. V. R. Dilshan, I. Weerathunga, and D. Kasthurirathna, "Sinhala sign language translation through immersive 3D avatars and adaptive learning," in *Proc. 2023 IEEE 8th Int. Conf. Adv. Comput. (ICAC)*, Dec. 2023, pp. 1–6.

[32] M. Priyankara, A. Gunasekara, and K. Ilmini, "Sign language translation techniques using artificial intelligence for the hearing impaired community in Sri Lanka: A review," in *Proc. 2023 IEEE Int. Conf. Artif. Intell. (SLAI-ICAI)*, Nov. 2023, pp. 1–6.

[33] S. Jayasena. (2009, Nov.) Sinhala text to speech system. [Online]. Available: https://www.academia.edu/33501396

[34] Y. Bestgen, "Comparing lexical bundles across corpora of different sizes: The Zipfian problem," *J. Quant. Linguist.*, vol. 27, no. 3, pp. 272–290, Feb. 2019.

[35] (2019) Acoustic modelling for under-resourced languages. [Online]. Available: https://core.ac.uk/display/1000014983

[36] S. Coperahewa, "The language planning situation in Sri Lanka," *Curr. Issues Lang. Plann.*, vol. 10, no. 1, pp. 69–150, Feb. 2009.

[37] H. H. S. N. Haputhanthri, H. M. N. Tennakoon, M. A. S. M. Wijesekara, B. H. R. Pushpananda, and H. N. D. Thilini, "Multi-modal deep learning approach to improve sentence level Sinhala sign language recognition," *Int. J. Adv. ICT Emerg. Reg.*, vol. 16, no. 2, pp. 21–30, Dec. 2023.

[38] D. T. D. M. Dahanayaka, B. G. D. A. Madhusanka, and I. U. Atthanayake, "A multi-modular approach for sign language and speech recognition for deaf-mute people," *Eng. J. Inst. Eng., Sri Lanka*, vol. 54, no. 4, p. 97, Dec. 2021.

[39] T. Wijesinghe, "Teaching-learning of Sinhala literacy skills in Deaf primary schools in Sri Lanka: A mixed methods inquiry," Ph.D. dissertation, University of Cambridge, Cambridge, UK, Apr. 2024.

[40] N. K. T. Navodya and et al., "Real-time motion detection for language translation and literacy enhancement in hearing-impaired children," in *Proc. 2024 IEEE Int. Conf. Inf. Technol. Res. (ICITR)*, Dec. 2024, pp. 1–6.

[41] N. de Silva. (2024, Jan.) Survey on publicly available Sinhala natural language processing tools and research. [Online]. Available: https://arxiv.org/abs/1906.02358

[42] K. Peffers, M. Rothenberger, T. Tuunanen, and S. Chatterjee, *Introduction to Design Science Research*, ser. Lecture Notes in Computer Science. Springer, 2012, vol. 7286, pp. 1–10.

[43] Google Fonts. (2020) Noto sans: A font family for a unified and inclusive digital world. [Online]. Available: https://fonts.google.com/noto/specimen/Noto+Sans

[44] J. Kim, S. Lee, and J. Lee, "The effects of text alignment on reading comprehension: A comparison between justified, left-aligned, and centered text," *J. Usability Stud.*, vol. 12, no. 1, pp. 1–10, 2017. [Online]. Available: https://uxpajournal.org/effects-text-alignment-reading-comprehension/

[45] W. Lidwell, K. Holden, and J. Butler, *Universal Principles of Design*, 2nd ed. Rockport, MA, USA: Rockport Publishers, 2010. [Online]. Available: https://www.rockpub.com/books/universal-principles-of-design

[46] D. Adiwardana and et al. (2020, Jan.) Towards a human-like open-domain chatbot. [Online]. Available: https://arxiv.org/abs/2001.09977

[47] S. Roller *et al.*, "Recipes for building an open-domain chatbot," Apr. 2021. [Online]. Available: https://aclanthology.org/2021.eacl-main.24

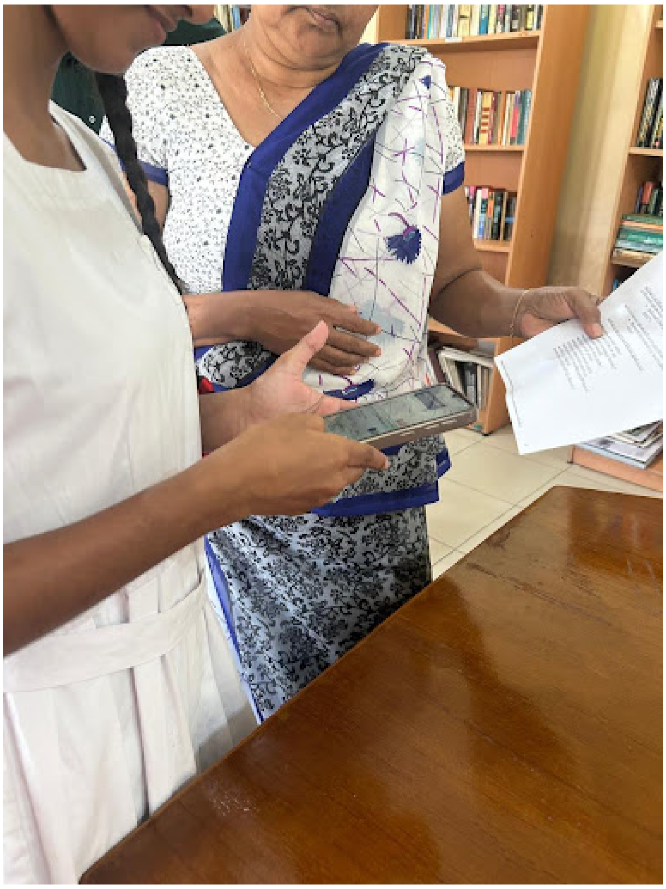
[48] Y. Graham, T. Baldwin, A. Moffat, and J. Zobel, "Can machine translation systems be evaluated by the crowd alone," *Natural Language Engineering*, vol. 23, no. 1, pp. 3–30, Sep. 2015.

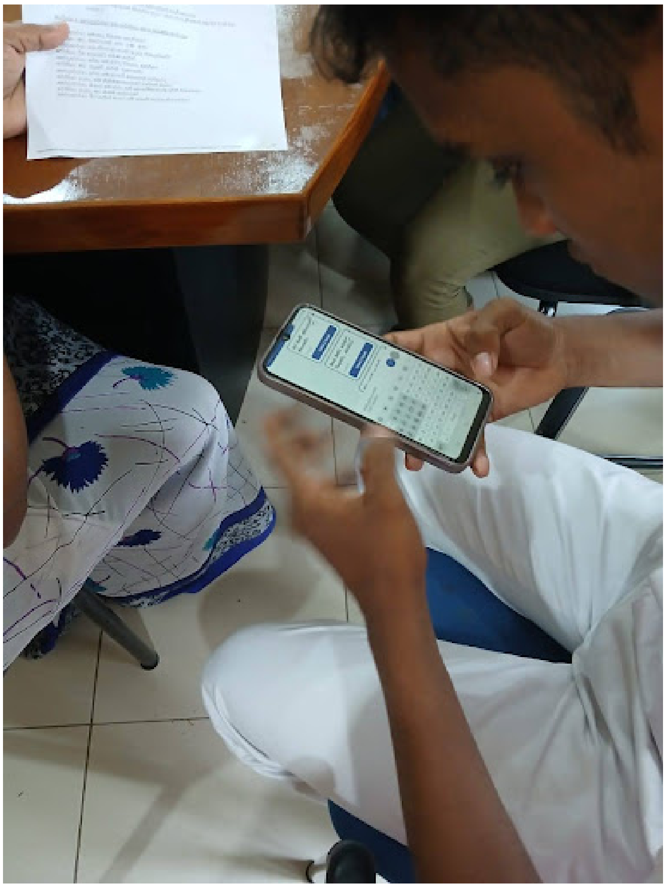
[49] A. Shvartzon, A. Azaria, S. Kraus, C. Goldman, J. Meyer, and O. Tsimhoni, "Personalized alert agent for optimal user performance," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, Feb. 2016.

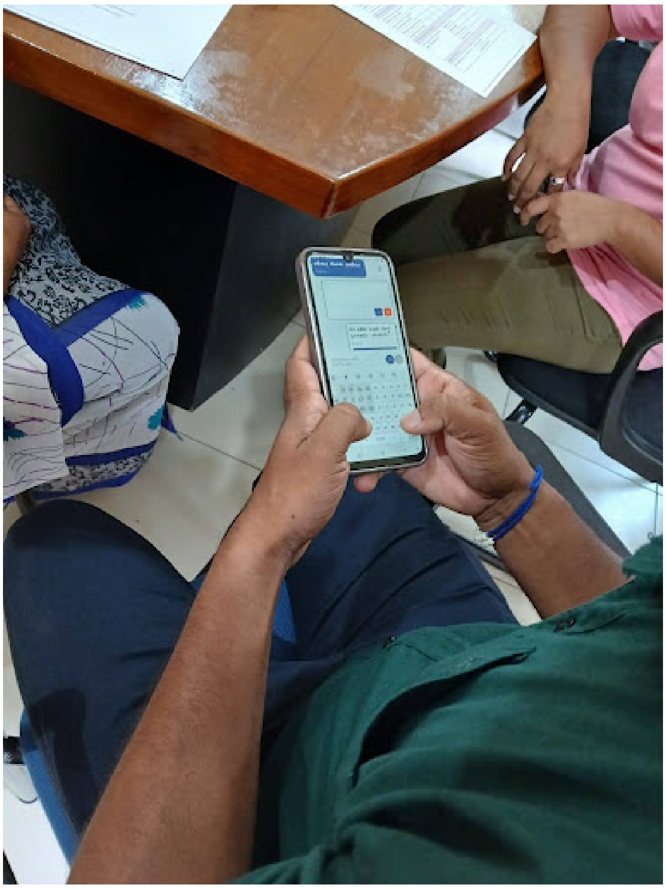
[50] L. Laranjo *et al.*, "Conversational agents in healthcare: A systematic review," *Journal of the American Medical Informatics Association*, vol. 25, no. 9, Jul. 2018.

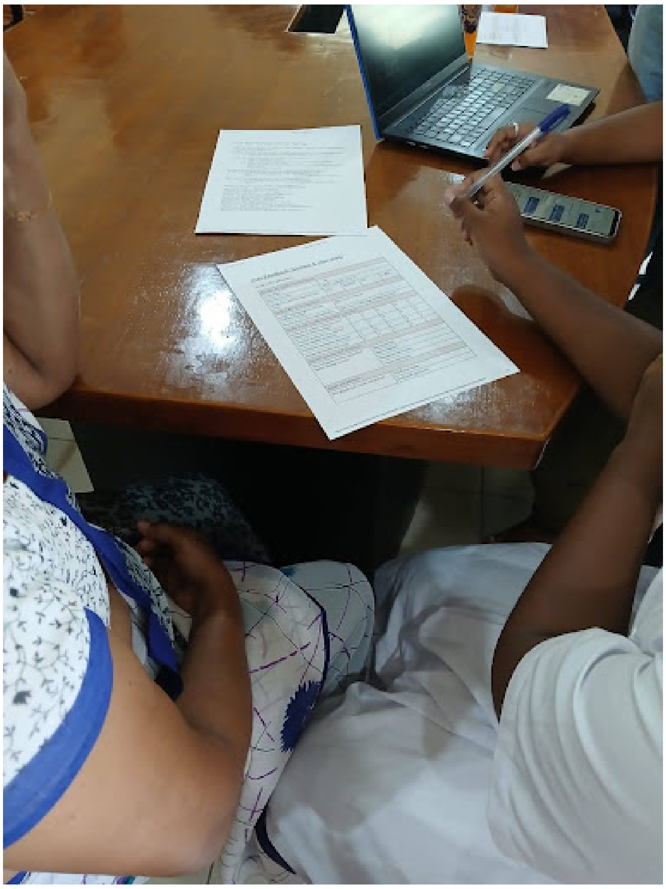
Appendices

.1 Interviews with Deaf and hard of hearing individuals

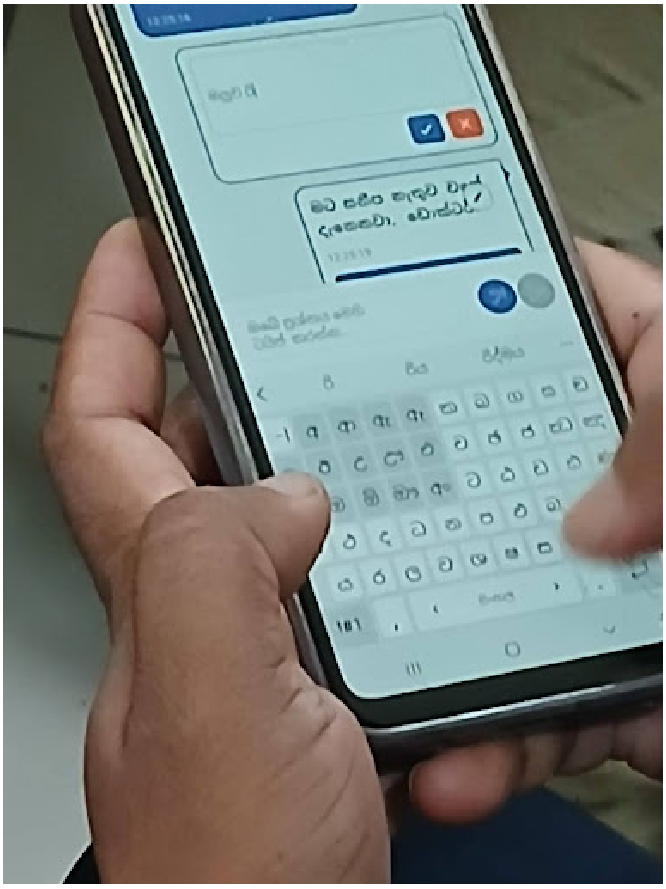









සවන් සහන
சவன் சஹன
SAWAN SAHANA
DR. REIJNTJES SCHOOL FOR THE DEAF.



.2 Conversation conducted between the simulated doctor and the patient

<u>කාර්යය 1: වෛද්‍යවරයා කතා කරයි → රෝගියා ප්‍රතිචාරය තෝරා ගනී/සංස්කරණය කරයි</u>
වෛද්‍යවරයාගේ (ශ්‍රවණාබාධිත නොවන පුද්ගලයින්) කාර්යභාරය:

- ප්‍රශ්නයක් කතා කරන්න (උදා: "ඔයාගේ රෝග ලක්ෂණ මොනවාද?").

රෝගියාගේ (ශ්‍රවණාබාධ (DHH) සහිත පුද්ගලයා) කාර්යභාරය:

- ප්‍රතිචාර නිරීක්ෂණය කරන්න.
- ප්‍රතිචාර 3 න් එකක් තෝරන්න (උදා: "මට හිසරදයක් සහ උණ තිබේ").
- දුෂ්කර වචන සඳහා රූප බලන්න.
- අවශ්‍ය නම් ප්‍රතිචාරය සංස්කරණය කරන්න (උදා: "මට දවස් දෙකකට කලින් උණ හැදුනා.").
- ප්‍රතිචාරය තහවුරු කරන්න

<u>කාර්යය 2: රෝගියා ප්‍රශ්නයක් ටයිප් කරයි → වෛද්‍යවරයා පිළිතුරු දෙයි</u>
රෝගියාගේ (ශ්‍රවණාබාධ (DHH) සහිත පුද්ගලයා) කාර්යභාරය:

- ප්‍රශ්නයක් ටයිප් කරන්න (උදා: "මම මේ බෙහෙත කොපමණ කාලයක් ගත යුතුද?").

වෛද්‍යවරයාගේ (ශ්‍රවණාබාධිත නොවන පුද්ගලයින්) කාර්යභාරය:

- සිංහලෙන් පිළිතුරක් කියන්න (උදා: "ආහාර ගැනීමෙන් පසු දින 5 ක් එය ගන්න").

<u>කාර්යය 3: වෛද්‍යවරයා සහ රෝගියා අතර අඛණ්ඩ සංවාදය</u>

වෛද්‍යවරයා: මොකක්ද තියෙන අසනීපය?
රෝගියා: මට බඩේ අමාරුවක් තියෙනවා
වෛද්‍යවරයා: කොච්චර කාලයක් ඉඳලද තියෙන්නේ?
රෝගියා: මාසයක් විතර වෙනවා
වෛද්‍යවරයා: වෙන මොනවද තියෙන අසනීප?
රෝගියා: මට දියවැඩියාව තියෙනවා
වෛද්‍යවරයා: වෙන මොනාහරි බෙහෙත් ගන්නවද
රෝගියා: ඔව් විටමින් පෙති ටිකක් ගන්නවා
වෛද්‍යවරයා: ඔයාට මොනවා හරි අසාත්මිකතා/ඇලර්ජි තියෙනවා.
රෝගියා: නැහැ, ඔව් මට දූවිලි වලට අසාත්මිකතාවයක් තියෙනවා
වෛද්‍යවරයා: මීට කලින් ඔයාට මේ ලෙඩේ හැදිලා තියෙනවද?
රෝගියා: නැහැ මට මේක පලවෙනි වතාවට

User Feedback (Surveys & Interviews)

යෙදුම් භාවිත සමීක්ෂණය

පුද්ගලික තොරතුරු (Personal Details)					
වයස (Age)	18				
ලිංගභේදය (Sex)	පිරිමි / ගැහැණු / වෙනත්				
ශ්‍රවණාබාධ මට්ටම (Hearing Loss Level)	මෘදු	මධ්‍යස්ථ ✓	දැඩි	ගැඹුරු	
පළමු හැඟීම (First Impression)					
යෙදුම් අතුරුමුහුණත තේරුම් ගැනීමට පහසුද?	ඔව් / නැත / තරමක්				
ඔබගේ භාෂා සහාය ප්‍රමාණවත්ද?	ඔව් / නැත / වැඩිදියුණු කිරීමක් අවශ්‍යයි				
කාර්ය සාධන ඇගයීම (Performance Evaluation)					
කථනයෙන් පෙළට නිරවද්‍යතාවය (Speech-to-Text accuracy)	1	2	3	4 ✓	5
පෙළෙන් කථනයට පැහැදිලි බව (Text-to-Speech accuracy)	1	2	3	4 ✓	5
(Response Generation accuracy)	1	2	3 ✓	4	5
(Image Generation accuracy)	1	2	3	4 ✓	5
වඩාත්ම ප්‍රයෝජනවත් විශේෂාංග					
ඔබ වැඩිපුරම භාවිතා කරන්නේ කුමන විශේෂාංගයද?	Speech-to-Text				
	Text-to-Speech				
	Response Generation				
	Image Generation ✓				
මුහුණු දත් අභිමතය					
ඔබ මුහුණු දත් ගැටළු කෙසේද?	පරිවර්තන දක්ෂ				
	ස්වයංක්‍රීය ප්‍රතිචාරය ✓				

	UI දුෂ්කරතා	
	වෙනත්	[illegible]
		[illegible]
යෙදුම වැඩිදියුණු කිරීම		
ඔබට යෙදුම වැඩිදියුණු කළ හැක්කේ කෙසේද?	වඩා හොඳ පරිවර්තනය	
	වේගවත් සැකසුම්	
	තවත් සිංහල සහාය	
	වෙනත්	[illegible]
අවසාන ප්‍රතිපෝෂණය (FeedBack)		
මෙම යෙදුම එදිනෙදා ජීවිතයේදී ප්‍රයෝජනවත්ද?	ඔව් / නැත	
ඔබ එය දිගටම භාවිතා කරනවාද?	ඔව් / නැත [illegible]	

සමස්ත තෘප්තිය (Overall satisfaction)	1	2	3	4 ✓	5

.3 TTS Evaluation Results

	Original Text	What did you hear?	Rating
	[illegible]	[illegible]	5
	[illegible]	[illegible]	2
Listener 1	[illegible]	[illegible]	5
	[illegible]	[illegible]	3
	[illegible]	[illegible]	5
	[illegible]	[illegible]	5
	[illegible]	[illegible]	5
	[illegible]	[illegible]	3
Listener 2	[illegible]	[illegible]	4
	[illegible]	[illegible]	1
	[illegible]	[illegible]	3
	[illegible]	[illegible]	5
	[illegible]	[illegible]	5
	[illegible]	[illegible]	2
Listener 3	[illegible]	[illegible]	2
	[illegible]	[illegible]	3
	[illegible]	[illegible]	5
	[illegible]	[illegible]	3
	[illegible]	[illegible]	3
	[illegible]	[illegible]	2
Listener 4	[illegible]	[illegible]	3
	[illegible]	[illegible]	3
	[illegible]	[illegible]	4
	[illegible]	[illegible]	2
	[illegible]	[illegible]	4
	[illegible]	[illegible]	4
Listener 5	[illegible]	[illegible]	4
	[illegible]	[illegible]	3
	[illegible]	[illegible]	4
	[illegible]	[illegible]	3
	[illegible]	[illegible]	5
	[illegible]	[illegible]	3
Listener 6	[illegible]	[illegible]	3
	[illegible]	[illegible]	3
	[illegible]	[illegible]	5
	[illegible]	[illegible]	2