ASD.AI: SINHALA DIALOGUE MANAGEMENT TOOL TO SCREEN KIDS WITH AUTISM SPECTRUM DISOREDR

Sampath G.A.D.M

(IT16061880)

BSc. (Hons) Degree in Information Technology Specializing in Information Technology

Department of Information Technology

Srilanka Institute of Information Technology

Srilanka

October 2021

ASD.AI: SINHALA SPEECH SYNTHESIS SYSTEM COMPONENT

Sampath G.A.D.M

(IT16061880)

Dissertation submitted in partial fulfillment of the requirements for the Degree Bachelor of Science Special (Honors) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

DECLARATION OF THE CANDIDATE & SUPERVISOR

Following the signature and date, the individual should make the following declaration. After consulting with the supervisors, a candidate might request a prohibition for a certain dissertation for a specific work for a specific period of time or indefinitely. An embargo like this might override the dissertation's own declaration.

“I proclaim that this is my own effort, and that this dissertation does not incorporate without affirmation any material previously submitted to a degree or diploma or Diploma at any other University or institute of higher learning, and that it does not encompass any subsequent iteration previously published or written by another individual to the best of knowledge and belief, except where affirmation is made in the text.

In addition, I grant the Sri Lanka Institute of Information Technology the indistinguishable right to procreate and disseminate my thesis in whole or in part in print, electronic, or other media. I maintain the right to incorporate this information into future works in whole or in part.

……. ……

Signature Date

The following statement should be signed by the dissertation's supervisors.

During my guidance, the aforesaid applicant conducted research for a bachelor's degree thesis.

…….. ……..

Signature of the supervisor: Date

ABSTRACT

The term "autism spectrum disorder" refers to a collection of developmental problems characterized by difficulties with social interaction, communication, and conduct. It is marked by a lack of social communication and the occurrence of repeated or strange conduct. The prevalence of ASD has risen dramatically in recent years. In the United States, for example, the Centers for Disease Control (CDC) reports that one out of every 68 children has been diagnosed with ASD. Many physicians have started testing for autism in the last decade. The goal of the autism test is to find the early symptoms of autism. A diagnostic tool, such as a modified checklist for detecting autism in children, is usually used to diagnose autism spectrum disorder. However, how behavioral symptoms of autism spectrum disorder are interpreted differs by culture: in Sri Lanka, the updated checklist for diagnosing autism in children has an accuracy of up to 25%. As a result, a sensitive cultural screening method for autism assessment has become a necessity. . A shortage of mental health specialists exists in low- and middle-income nations, which is a major impediment to early identification of autism spectrum condition. Early diagnosis of autism spectrum disease allows for intervention prior to the emergence of abnormal behavioral and cognitive characteristics. This article offers a culturally appropriate autism spectrum disorder screening tool. For the very first time, the proposed application requires a clinically validated symptom checklist to identify and detect autism spectrum disorders in low- and middle-income nations like Sri Lanka.

Hidden Markov Model (HMM) speech synthesis is one of these techniques that has recently been proved to be exceptionally effective in synthesizing acceptable speech. This review also compares and contrasts these methods with the more traditional unit selection synthesis method, which has dominated speech synthesis for the past decade. We will discuss the benefits and drawbacks of statistical parametric synthesis, as well as where, in our opinion, major developments may occur shortly. This article explores several text-to-speech synthesizer approaches for obtaining comprehensible and natural output, with the cascading formant method being used to create the voice synthesizer. When you type text, text-to-speech conversion is based on the development of appropriate sound. Speech synthesis is employed in a variety of settings, including medical, telecommunications, and other fields. Hidden Markov Model (HMM), Formant, Articulated, and Concatenative are some of the voice synthesis approaches that have been utilized to transform text to speech for comprehensible and natural inference.

Keywords – Hidden Markov Model (HMM), Text-to-speech (TTS)

ACKNOWLEDGMENT

The study presented in this article was done as part of the fourth-year research project in our CDAP program. This initiative is the outcome of the institution's arduous efforts, as well as the support, assistance, and guidance provided by many others. Our team would like to express its gratitude to everyone who assisted us in completing this difficult task.

Professor Koliya Pulasinghe, the task supervisor, deserves our gratitude for his constant support, advice, and aid in the development of this project, particularly for the scholarly and thought-provoking discussions.

Professor Koliya Pulasinghe and MS Vijani Priyanwadha owe our heartfelt gratitude for guiding us through this process. The work mentioned in this article was done as part of the fourth-year research project for our CDAP program. This achievement is the result of the group's hard work, as well as the inspiration, encouragement, and guidance supplied by a large number of individuals. Our crew would like to express its heartfelt gratitude to everyone who assisted us in completing this challenging assignment.

MS Vijani Priyanwadha, our Co-Supervisor, is also thanked for all of her assistance and suggestions in the development of the research project, particularly in our documentation efforts. Professor Koliya Pulasinghe deserves our gratitude for putting up with our investigation proposal, assisting us in selecting a suitable manager at the start of our investigation, and providing us with all of the advice and support we required during the project. Furthermore, our team is grateful for the assistance, suggestions, and advice received from our colleagues at the Sri Lanka Institute of Information Technology, as well as their enormous help and direction in making this item a success. Any remaining gatherings' assistance is also acknowledged. Finally, we'd want to thank everyone else who has supported us in various ways and encouraged us to make this a success but whose names aren't mentioned here.

TABLE OF CONTENT

Contents

TABLE OF CONTENT iv
LIST OF FIGURES vi
LIST OF ABBREVIATIONS viii
1. INTRODUCTION 9
1.1 Background Literature 9
1.1.1 MSSAP 14
1.1.2 eSPEAK 15
1.1.3 Festival 15
1.1.4 Crazy Talk 15
1.1.6 MBROLA 15
1.1.7 UCSC Sinhala TTS 16
1.2 Research Gap 17
2. Research Problem 19
3. Main Objectives 22
4. RESEARCH METHODOLOGY 25
4.1 Requirement gathering and analysis 25
4.2 System Overview 26
4.3 System Process 26
4.4 Phonotactics of Sinhala Language 29
4.4.1 Diphthongs 29
4.4.2 Nasal Words in Sinhala 30
4.5 Sinhala Prosodic Features 31
4.5.1 Stress 31
4.5.2 Pitch 31
4.5.3 Intonations 32
4.6 How to Build a Synthetic Voice from a Text using the HMM? 32
4.6.1 Training Stage of HMM 33
4.6.2 Synthetic Stage of HMM 35
4.6.3 Speaker Conversion 35
4.7 Commercialization of the Product 36
4.7.1 Customer Relationships 39

4.8 Tools and Technologies…… 40
5. Testing & Implementation 43
5.1 Types of Testing…… 44
5.2 Implementation…….. 44
5.2.1 Proof of Implementation 45
5.3 Test Case Scenarios…… 51
5.3.1Test Scenarios for Login function 51
5.3.2 Test Scenarios for Speech Signal Processing function 54
5.4 Types of Testing…… 54
6. RESULTS & DISCUSSION 58
6.1 Results……58
6.2 Research findings and Discussion 58
6.3 Conclusions……59
7. References 60
Appendices 63
Appendix A- Plagiarism Content…… 63

LIST OF FIGURES

Figure 1. 1 The expansion and the milestones of speech synthesis 14
Figure 1. 2 Speech Generation Function 17

Figure 4. 1 Component Overview Diagram 26
Figure 4. 2 Speech Process 27
Figure 4. 3 Speech Generation from a random input 28
Figure 4. 4 Speech Generation Methodically 29
Figure 4. 5 Diphthongs in Sinhala Language 29
Figure 4. 6 Speech Signal Converted to recognizable word 33
Figure 4. 7 Synthesis Speech System 34
Figure 4. 8 Speaker Conversion 35
Figure 4. 9 ASD identification through consultation 37
Figure 4. 10 Python Logo 40
Figure 4. 11 RASA Framework Logo 41
Figure 4. 12 Google Collab Logo 41
Figure 4. 13 Visual Studio Code Logo 41
Figure 4. 14 Android Studio Logo 43

Figure 5.4. 1 Functional Testing 55
Figure 5.4. 2 Nonfunctional Testing 56

Figure 5.2. 1: Pretrained data for speech synthesis 46
Figure 5.2. 2: Download the pretrained data model 46
Figure 5.2. 3: Depiction of the downloaded pretrained data set 47
Figure 5.2. 4: Synthesized speech text depiction 47
Figure 5.2. 5: Synthesizing a text illustration 48
Figure 5.2. 6 :Text conversion to speech illustration 48

LIST OF TABLES

Table 4. 1 Personalized consonants in Sinhala Language 31

Table 4. 1 Personalized consonants in Sinhala Language 31

Table 5.3. 1 Test Case for Login Scenario..... 51
Table 5.3. 2 Test case for Unsuccessful login Admin Side 52
Table 5.3. 3 Test Case for unsuccessful login- User side..... 53
Table 5.3. 4 Test Case for speech signal successful input..... 54
Table 5.3. 5 Test case for speech signal unsuccessful input..... 54

Table 5.4. 1 Functional Testing 55
Table 5.4. 2 Nonfunctional Testing 57

LIST OF ABBREVIATIONS

Abbreviations	Description
HMM	Hidden Markov Model
TTS	Text to Speech
CDC	Center for Disease Control
ASD	Autism Spectrum Disorder
IVR	Interactive Voice Recognition
OCR	Optical Character Recognition
PC	Personal Computer
AI	Artificial Intelligence
GPU	Graphical Processing Unit
IDE	Integrated Development Environment
NLG	Natural Language Generator
ICT	Information Communicational Technology

1. INTRODUCTION

1.1 Background Literature

There has been a noticeable surge in research focusing on discovering biological and behavioral indicators to aid in the early detection of Autism Spectrum Disorders in recent years (ASD). ASD is a collection of disorders marked by deficits in social, linguistic, and communication, as well as repetitive stereotyped behaviors. Early diagnosis is essential for improved treatment response and reduced caregiver burden. Autism is known to present itself in a variety of ways in children's and adults' communication. Echolalia, out-of-context wording, and pronoun and reversal of roles are all common language anomalies . [1] However, there are various different subtypes of language skills in autism. As a result, linguistic-based indicators may not be effective for ASD identification. Paralinguistic indicators, on the other hand, seem to be ideally equipped for automatic identification than aberrant prosody, which has also been described as a fundamental hallmark of ASD. [2] Suprasegmental acoustic aspects related to articulation, loudness, tone, and cadence have shown promising results for children's speech. These acoustic elements have also been effectively used in speech-based engagement systems to improve the social skills of children with ASD. Machine learning approaches based on auditory and prosodic feature sets have been investigated to diagnose autism automatically . [3] While studies suggest that high levels of reliability can be reached for tasks such as distinguishing positive phase children from kids with ASC, such systems' performance has been evaluated on very short datasets, which could lead to potential confounds . The minimal number of currently available ASD -related datasets is a major roadblock to developing robust models that are dependable enough for clinical use . Nonetheless, the vision challenged population in Sri Lanka has a hard time communicating with technology because an appropriate tool is not readily available. Through this research thesis we are going to talk in depth about what is a TTS system and its historical background. [4]

Another important discussion is how to treat the children and adults that are suffering from ASD? In this venture If a doctor suspects that a child has ASD or another developmental problem, the kid will be referred to a range of specialists, including a speech-language pathologist.[5] A voice, speech, and language therapist is a health practitioner who specializes in treating people who have problems with their voices, speech, and language. The speech-language pathologist will do a thorough assessment of the child's communication abilities before recommending a treatment plan. A referral for a hearing test may also be made by the speech-language pathologist to ensure that the child's hearing is normal. It is critical to help children with Autism spectrum disorder improve their communication skills and the ability for them to attain their maximum capabilities. [6] There are a range of methods, but the most effective treatment begins in preschool and is tailored to the child's age and preferences. It should target the child's conduct as well as his or her communication abilities, as well as provide regular reinforcement of favorable behaviors. Highly structured, specialized programs work successfully for most children with ASD. The therapy program should include parents or primary caregivers, as well as other family and friends, so that it becomes a part of the child's everyday life. [7]

Developing speech and language skills is a reasonable therapy aim for some younger children with Autism spectrum disorder. By bringing awareness to a child's language acquisition early on, parents and caregivers can improve his or her chances of achieving this goal. Children develop pre-language skills before they start to use words, much as infants learn to crawl before walking. [8] Eye contact, gestures, body movements, mimicking, and babbling and other tones of voice are some of the techniques they use to communicate. A speech-language pathologist can assess and treat children who lack these abilities to prevent further developmental issues.[9]

Interaction training provides fundamental speech and language abilities, such as single words and phrases, to somewhat older children with ASD. Advanced training focuses on how language can be used to accomplish a goal, such as how to maintain a conversation with another person, which includes remaining on subject and taking turns responding. Verbal language and communication skills may never improve in some children with ASD. [10] Learning to communicate with gestures, such as gestures, may

be the objective for these youngsters. Others may seek to communicate via a symbol system in which images are utilized to express ideas. Symbolic systems can range from simple graphic boards or cards to complex electrical gadgets that create speech by pressing buttons to depict everyday objects or actions. [11]

Speech is the primary medium via which we communicate with one another in our daily lives. When it comes to engaging with machines, nevertheless, the majority of interaction is now accomplished through reading the computer screen, in addition to viewing and executing actions. It entails a lot of time-consuming activities such as visiting the internet, reading emails, eBooks, journal articles, and so on. Nonetheless, the visual impairments community in Sri Lanka has a hard time communicating with laptops because an appropriate tool is not readily available.[12]

This research presents an efficient tool for Text-To-Speech conversion that accommodates speech in native language as a viable answer to this challenge. In order to get a brief understanding about what is a “Text to Speech System” let us explore the term a bit more. When something is displayed on a display or handwritten, not everyone can read it. This could be due to the fact that the individual is partially sighted or illiterate. These persons can be helped by employing a Text-to-Voice (TTS) System to generate speech for the given text rather than printing or displaying it. A Text-To-Speech (TTS) system takes written text as input (from a web page, text editor, clipboard, etc.) and converts it to an audible format, allowing you to hear what's written. It recognizes what is presented on the monitor and reads it aloud.[13]

A TTS program allows you to listen to computerized text rather than comprehending it. That means you can attend to your emails or read eBooks while doing anything else, so improving efficiency. TTS can be used to resolve the literacy barrier of the general public, increase the potentials of improved man-machine interaction through on-line daily paper reading from the internet, and enhance other information systems such as gaining knowledge guides for schoolchildren, IVR (Interactive Voice Recognition) systems, and automated weather forecasting, among other things. Text-To-Speech is compatible with practically all personal digital devices, such as PCs, cellphones, and tablets. All types of text files, including Word and Pages documents, can be read aloud. Even web pages on the internet can be read aloud. TTS uses a computer-generated

voice, and the reading pace can normally be adjusted. [14]The quality of the voices varies, but some of them sound human. As words are read aloud, several Text-to-Speech programs highlight them. This allows individuals to see text while also gearing it. Optical character recognition is a technique used by several TTS programs (OCR). TTS programs can read content aloud from photos thanks to optical character recognition (OCR). You could, for example, photograph a street sign and have the text on the sign converted to audio. Text-to-Speech software comes in a variety of forms. The users can use a variety of tools based on the technology. [15]

- Many gadgets feature built-in text-to-speech (TTS) capabilities. Chrome is compatible with computers and laptops, as well as smartphones and digital devices. The user does not need to buy any specific apps or software to use this TTS.
- TTS tools on the web: Some webpages have on-site TTS tools. For example, to have this webpage start reading, the user can utilize the website's "Reading Assist" option, which is accessible in the lower left of the image. Dyslexics may also be eligible for a free Book Collaborative efforts, which includes digital books that may be viewed with TTS. [16](Understood founding partner Beneath runs a book share program.) TTS tools are also available for free on the internet.
- Text-to-speech software packages include the following: For personal computers, there are also several literacy software programs. Many of these apps have TTS in combination to other reading and writing tools. Kurzweil 3000, Clamored, and Read Write are other examples. TTS is also included in Microsoft's Immersive Reader product. It can be found in Microsoft Office apps such as OneNote and Word. More software for youngsters with reading problems can be found here. [17]

Prior to the start of the “ASD.AI” project, a basic research was done to become acquainted with TTS systems and to obtain information on current TTS systems. After that, a thorough analysis of relevant literature was conducted in the areas of Sinhala language and its characteristics, Festival and Festvox, generic TTS architecture, constructing new synthetic voices, Festival and Windows integration, and how to improve current voices. Understanding how contemporary systems work and how

they've evolved into their modern version requires a historical analysis. History of Speech Synthesizing will cover the evolution of speech data from mechanical synthesizing to today's high-quality synthesizers, as well as some key milestones in synthesis-related approaches. [18]

Wolfgang von Kempelen invented the "Acoustic-Mechanical Speech Machine" in 1791, which produced individual and groups of sounds. In his writings, he presented his research on speech production and tests with his speech machine. The main components of his device were a pressure chamber for the lungs, a buzzing reed to act as vocal cords, and a leather tube for the vocal tract movement, and he was able to make varied vowel sounds by altering the curvature of the material tube. For phonation sounds, a replica of the vocal folds with a hinged tongue and adjustable lips was used, while syllables were made by four discrete constricted passages manipulated by fingers.

Charles Wheatstone built a variant of Kempelen's talking apparatus in the mid-nineteenth century that could generate vowels, attuned, some sound combinations, and even complete words. Consonants, including nasals, were formed with turbulence through a suitable passage with reed-off and vowels were produced with vibrating reed with all passageways closed. [19]

Stewart introduced the first entirely electrical synthesis apparatus in 1922. There was a buzzer for stimulation, as well as resonant frequency circuits to replicate the vocal tract's acoustic resonances. Single static phonetic symbols with two lowest formants were produced by this machine. It couldn't do consonants or linked utterances, though. Winger created a synthesizer that was comparable to this one. This gadget was made up of four electrical resonators connected in series, as well as a buzzing source. [20]

Vowel spectra were created by combining the four outcomes of resonators at the appropriate amplitude and frequency. Obata and Teshima [28], two Japanese scholars, discovered the third formant in vowels in 1932. For comprehensible synthetic speech, the first three formants are usually regarded sufficient. [21]

The VODER (Voice Operating Demonstrator), unveiled by Homer Dudley at the 1939 New York World's Fair, was the first gadget that might be termed a speech generator. The VODER was influenced by the VOCODER (Voice CODER) built at Bell

Laboratories in the mid-1930s, which was primarily designed for communication. The VOCODER was created as a voice transmitting device to replace low-band telephones.[22] It processed wideband speech, changed it into slowly fluctuating control signals, delivered them via a low-band phone line, and then converted those signals back to the correct speech. Noriko Umeda and his co - workers developed the very first full text-to-speech program for English in the Electro Technical Laboratory in Japan in 1968. A syntactic evaluation module with some complex heuristics was incorporated in the synthesis, which was based on an articulatory model. Despite being understandable, the system is monotonous. [23]

The below figure shows the expansion and the milestones of speech synthesis over the years.

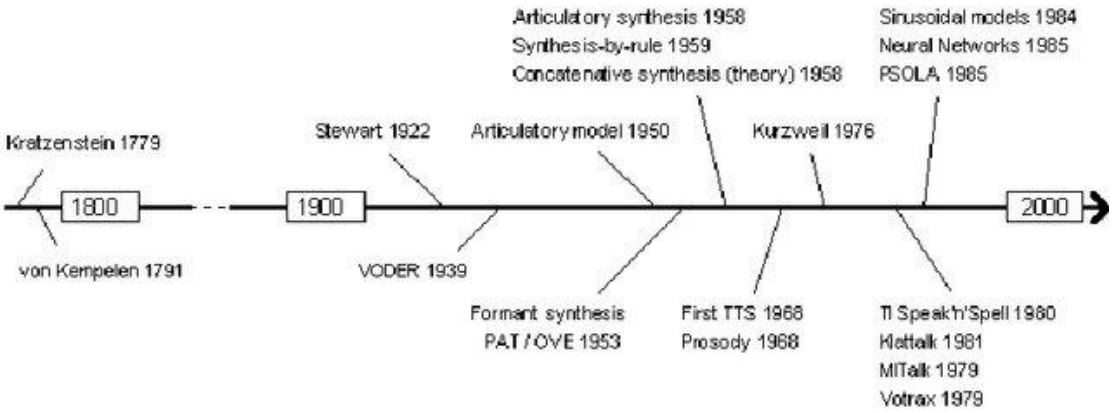


Figure 1. 1 The expansion and the milestones of speech synthesis

There are numerous TTS systems available for a variety of languages. The majority of them were created with the English language in mind. Because English is the most extensively used language in computer networks, English TTS systems have a distinct advantage over systems in other languages. As a result, in addition to existing Sinhala TTS systems, members of the "ASD.AI" project researched some of the most extensively used English TTS systems. [24]

1.1.1 MSSAP

There are numerous TTS systems available for a wide range of languages. The majority of them were created with the English language in mind. Because English is the most extensively used language in computer networks, English TTS systems have a distinct

advantage over systems in other languages. As a result, in addition to existing Sinhala TTS systems, members of the " ASD.AI " project researched some of the most extensively used English TTS systems. [25]

1.1.2 eSPEAK

The Free Software Foundation, Inc. developed eSpeak, an open software TTS system. It was first released in 1995 under the name talk and was designed for Acorn/RISC OS machines. This updated iteration, dubbed eSpeak, is a rewrite and update that includes a relaxation of the original capacity and processing power limits, as well as support for other languages. The quality of voice in eSpeak is found to be fairly low, resulting in a robotic speech production.[26]

1.1.3 Festival

Festival is a bilingual And multilingual Synthesis System developed by the University of Edinburgh's Centre for Speech Technology Research. It includes tools for creating TTS for a number of languages for which funds are provided, as well as general speech synthesis tools for other languages. The project's fundamental operations are written in C++, and these methods are encased in packages written in the Scheme programming language. Any changes to Festival for a new language can be made in Scheme rather than C++, in an area that enables for easy code update.[27]

1.1.4 Crazy Talk

Crazy Talk is a commercialized TTS application. There are male and female accents, as well as a variety of emotive imagery options. Crazy Talk allows the user to choose a voice and a sentiment, resulting in an acoustic signal with an emotive imagery. A person can also use his or her own picture as the presenter's image.

1.1.6 MBROLA

MBROLA is a voice synthesis based on the concatenation of diphones, developed by the TCTS (Théorie des Circuits et Traitement du Signal) Lab of the Faculty

Polytechnique de Mons (Belgium) in 1996. It takes a list of phonemes as information, along with syllable data (phoneme period and a piece - wise explanation of angle), and generates 16-bit (linear) speech specimens at the sampling frequency of the set of relevant database used (it is not a Text-To-Speech (TTS) synthesizer, because it does not recognize raw text as input).

1.1.7 UCSC Sinhala TTS

This is the only Sinhala TTS that is made accessible to the Sri Lankan people. The technology was developed as part of a UCSC study. It only has a male voice named "ucsc_sin_sdn." The effortlessness of the voice and correct utterances are also important flaws in UCSC TTS. The voice is a little robotic and lacks a lot of speed fluctuations. Furthermore, it misreads numerous commonly used words and mathematical operators.

Despite voice command systems such as festival, eSpeak, MS SAPI, Crazy Talk, and MBROLA are utilized, there are some significant variations between them. With the exception of MS SAPI and Crazy Talk, all three components are compatible with both Windows and Linux platforms. Only Windows systems are compatible with the MS SAPI TTS engine. As a result, we can utilize it to generate voices directly in Windows apps. Crazy Talk is a TTS system that is also available as business applications. It's also able to make facial emotions. This likewise only works in Windows-based settings. Another feature is that it can be used with browsers such as Internet Explorer and Mozilla Firefox. This engine demands more resources than other TTS engines.

MBROLA is not a comprehensive TTS synthesis platform, while festival and eSpeak are. The synthetic voice is generated using diphone files, and raw text is not accepted as input. To produce diphones, it requires additional instruments. Festival can accept raw text and generate diphones for the synthesizer on its own because it includes language modeling capabilities.

In addition, the MBROLA generated voice has a far greater vocal quality than the festival synthetic voice. It has a more natural sound than festival. Because these two methods are compatible, we can utilize festival to recognize phonemes and generate the

diphones that MBROLA needs to generate speech. This will aid in the creation of a voice that is more natural sounding.

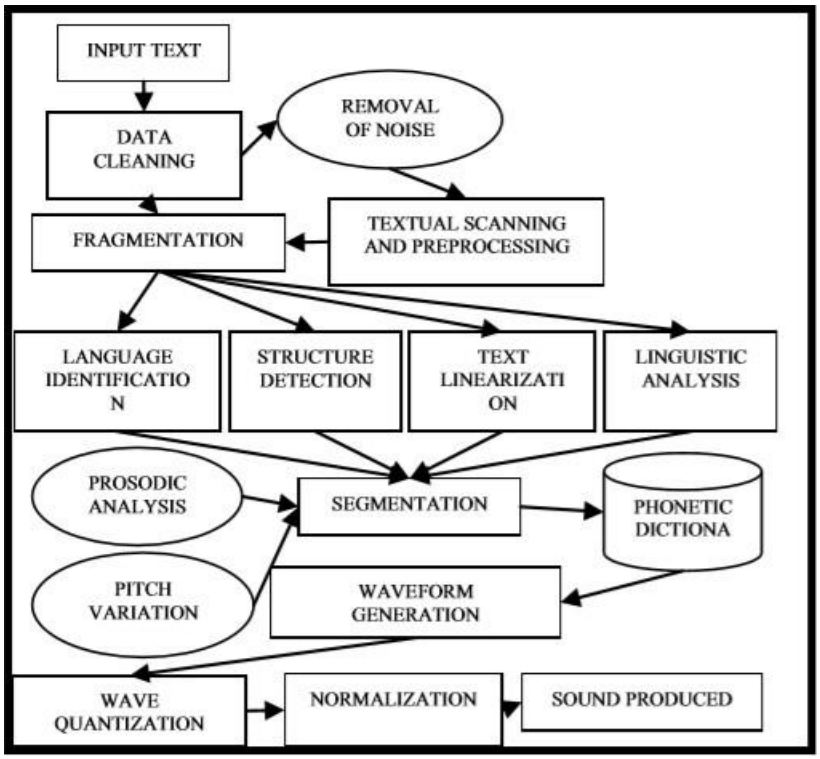


Figure 1. 2 Speech Generation Function

1.2 Research Gap

There is a research gap. Subject or area in which the capacity to draw a decision for an issue is hampered by a lack of or incomplete information. What is the connection between a research gap and research limitations? It is a gaping hole in the current research literature. It's the void that your research strategy fills. A complete literature survey and review is used to identify gaps. In relation to the potential solutions presented in this article, significant research has been undertaken in the following categories: speech recognition, interactive AI, and voice synthesis. Many of the studies looked at resulted in the product being able to do the following duties, such as the one proposed in this work. Thanks to advancements in speech recognition technology, the platform now has multilingual capability for real-time speech processing. The majority of systems, according to studies completed on each system, do not support speech to

text. Presently, the user of a speech synthesis system is confined to the platform's voices, which are typically provided to listeners immediately. After a trial run, it's difficult to locate a supportive platform to develop your voice for a more natural tone. We were unable to locate a Sinhala-based platform, despite the fact that the majority of these platforms cover languages spoken all over the world. [30]

The ability to handle the current state of a discussion has improved this type of investigation, replacing the old unstable dialog control into a state dialog. We were able to focus on deep learning-based neural network access for conversational control research thanks to advancements in GPU configuration and machine learning. This method has been used extensively in recent study in this field. In this field of study, the subject of supporting the Sinhalese language is still relevant.

The suggested platform, on the other hand, will primarily focus on Speech Synthesis. The platform will be a stand-alone solution that can be implemented locally and customized to fit the needs of the company. Each spoken word is formed by combining a set of vowel and consonant speech sound components in a phonetic manner. Speech synthesis is the process of creating artificial human speech. The several sorts of speech synthesis methods that can be used to improve text to speech are discussed. The mechanism is likely to acquire structured when both intelligibility and naturalness define the maximum to generate artificial speech sounds. The formant synthesis output speech is more likely to have a robotic and artificial sounding voice. Each uttered word is formed by combining a set of vowel and consonant spoken sound components in a phonetic manner. Voice Synthesis is the technique of transforming any arbitrary writing in any language into a matching speech sound unit. Text-To-Speech is another name for this (TTS). Many academics chose this issue for their studies because it is currently a hot topic in the world of information technology. However, this problem has been discussed in this field since the 18th century. However, scholars in this sector are more interested in the machine learning technique. Many different forms of study have been conducted in this topic since its inception, using a variety of languages. Because the input is unexpected when utilizing speech synthesis, researchers have turned to a context-independent technique to address the unpredictability of the input. Even people

can learn to pronounce unfamiliar phrases quickly by observing how other ones are pronounced. This is the foundation of this context-agnostic method.

2. Research Problem

Autism spectrum disorder (ASD) is a cognitive disease that can lead to major social, communicative, and physical difficulties. Children with Asd can have a spectrum of disorders, skills, and levels of disability, which is referred to as the "spectrum." ASD has a variety of effects on people, ranging from moderate to severe. Some concerns, such as difficulties with social contact, are shared by children with Asd, but there are variations in when symptoms begin, how severe they are, the variety of signs, and whether or not other issues are present. The clinical manifestations can alter over time.

Children with ASD's capacity to speak and use language is determined by their educational and spiritual development. Some children with ASD may be unable to interact through speech and language development, while others may have poor speaking abilities. Others may have extensive vocabularies and be able to speak in depth on specific topics. Many people have difficulty understanding the meaning and rhythm of words and sentences. They may also have trouble deciphering body language and the meanings of various speech tones. These issues, taken combined, have an impact on children with ASD's capacity to engage with others, particularly peers their own age.

Technology improves every day, making people's lives easier. However, there are several aspects that remain unexplored. ASD awareness is limited in LMICs like Sri Lanka due to cultural reasons. Due to a lack of resources, patients with ASD are frequently left untreated for long periods of time. To enhance clinical outcomes in small children with ASD, early detection and diagnosis are critical. Smart gadgets are a good platform for a computer-aided tool since they are widely available and widely used. Most existing screening instruments robotize standard screening agendas like M-CHAT R/F. Just the ASD.AI application installs a keen AI model to show up at a choice.

Speech synthesis is the technique of a machine creating pronounced words from written information. There are many different sorts of input that can be supplied. Synthetic speech systems make two basic types of pronunciation mistakes. Text-to-speech (TTS) systems frequently get it wrong when it comes to decoding words . Consider how

difficult this will be, particularly for uncommon words spelled in unusual ways or words with two possible homographs that are pronounced differently (for example, "put" the common verb and "put" the less commonly pronounced verb that has to do with golf). It is sometimes combined with other languages like as Sinhala, Tamil, and English. The technique that will be used to translate text to speech should be able to recognize many languages as a single input.

Modern TTS devices have excellent sound experience, but they have difficulty spelling rare words. Unnatural prosody is perhaps the most serious issue they have. The term "prosody" refers to the rhythm, tone, and other aspects of speech that span numerous words. Prosody is tough for TTS to master since it requires the computer to understand the idea of what it is saying. At the prosodic level, there's an unlimited variety of ways to say anything, unlike at the phonetic tier, where there is often just one way to spell a word in a specific dialect. Because it excels at mimicking the source speaker's prosody, speech-to-speech voice translation has a natural edge in prosody over TTS (and the source speaker, hopefully, does understand the text). TTS systems produce considerably less natural sounding prosody than Respeecher's technique. It provides resource developers with a limitless prosodic palette.

The most fundamental is homograph inconsistency, which occurs when two words have distinct meanings but the same written form. Syntactic ambiguity is also common. Misunderstanding of any kind adds to the difficulty of producing high-quality speech synthesis. Problems exist with names as well, because persons with the same first or last name can pronounce them differently and identifying such a name in a text is equally challenging. Confusion is frequently the outcome of a conflict between conflicting viewpoints. This tension can heighten a user's interest and involvement. In natural speech, ambiguity is frequently utilized to achieve a certain impact, such as irony. This refers to a phrase with a deliberate contradiction between apparent and intentional meaning. The employment of an opposing feeling to the spoken material, such as the words "What a great day" delivered with rage, is one way to generate irony. When meaning and emotion are contrasted in this way, a multifaceted image of the speaker emerges. It conveys more than the simple statement "What an awful day." The majority of current voice command systems produce neutral speech. The evaluation of

an unique set of emotional responses, such as terror, rage, happiness, and astonishment, is the subject of research on expressive speech synthesis that clearly communicates emotions. Speech synthesis relies heavily on text preprocessing. Full sentences must be formed from digits and numerals. In English, the digit 182 would be enlarged to one hundred and eighty-two, and 1692 would be extended to sixteen hundred and ninety-two (if year) or one thousand six hundred and ninety-two (if year) (if measure). Fractions and dates are difficult to work with. 5/19 can be written as five-nineteenth (if it's a fraction) or May nineteenth (if it's a date) (if date). When compared to others, the first three ordinals must be pronounced differently. 1^{st} should be spoken first, 5^{th} should be second, and 3^{rd} should be third. The same issue will arise with roman numbers. The third lecture should be pronounced as the third. Queen Issabelle II, on the other hand, should be pronounced Queen Issabelle the second. Abbreviations should also be extended into their complete form. However, there are several issues with the context. Kg is a unit of measurement that can be either kilograms or kilograms.

The next objective is to determine the proper pronunciation for various contexts in the text. Homographs are a type of word that causes the most difficulty with text-to-speech systems. Some words, such as "read," have many pronunciations and tenses, such as present tense and past participle. The most difficult challenge is determining proper intonation, stress, and length from printed material. Prosodic characteristics are what they're called. When pronouncing the written words, the speech should have the proper rhythmic, tone, and intensity. It's challenging to precisely time sentences or organize words into phrases. For example, the input text "Alex says North is a singer" can be spoken in two ways, yielding two alternative meanings: "Alex says North is a singer" or "Alex says North is a singer." North is a vocalist in the first phrase, and Alex is the singer in the latter.

Special characters and symbols such as the dot ".", question mark "?", and hash "#" are frequently ignored by online voice translators. Only a few prefabricated voices are usually available in their databases for synthesis. The use of modern software frequently results in an altered pronunciation of a text. Furthermore, the quantity of words in the input sentence that will be transformed into speech has a restriction.

3. Main Objectives

The major goal of the component is to create a fully functional Sinhala Text to Speech system that produces speech that sounds like a human voice while keeping the Sinhala language's unique prosodic qualities in order to assist the autistic kids in Srilanka for the development of their communicational skills. A voice will be included in the system, which is a significant need in the present localization software market. To develop a TTS system with the ability to maintain a real-time conversation with Autistic kids is another major objective of initiating the research component. With the implementation of the afore mentioned initiatives, the autistic kids in Srilanka will be benefited by the following ways;

1. Speech therapy can help autistic kids to communicate more effectively. Children with autism can increase their ability to develop relationships and operate in daily life as a result of this.
2. Understand verbal and nonverbal communication, as well as the objectives of certain kids , in a variety of situations.
3. Interact with the society in a more beneficial way and articulate the words well.
4. Exchange the ideas and develop the conversational ideas in beneficial way.
5. Kids can enjoy their normal day today life by playing, enjoying as the normal kids without any hindrance.

The system will be built on the Deep Voice 3 and Wave NET platforms, with assistance for English and Sinhalese languages. A module including an encoder, decoder, and transformer will be used to train the recurrent neural network. The encoder turns the text functions into an internal representation that it has learned. The transformer creates post-processed audio to make the voice more realistic and human, whereas the decoder interprets the remembered performances.

Another major objective following the component is to create a TTS system that can correctly pronounce the given text in terms of rhythm and melody so that the kids with autistic behavior will find it more attractive and they will direct more of their attention towards the system, and it will be way more convenient for them. When a given text or phrase is uttered in terms of rhythm and melody it is way easier for the kids to understand it and also this is a much better idea for the implementation with the autistic

kids specially when they have some physical impediments when compared to the normal children. Also, by initiating such a system the kids will also be treated as same as the other population in the country and they will not be treated separately.

To determine right pronouncing in various instances throughout the text, as well as correct intonation, stress, and duration is also another feature that was took into consideration when implementing such a system for the autistic kid population in Srilanka. If the system is able to capture the correct feelings and emotions, it will be very meaningful, and the attempt would also be a huge success and ultimately the kids will also be benefitted. We also identified some other specific objectives in addition to the afore mentioned ones, they are;

1. Do the text pre-processing task to overcome the problems discussed in research problem part

   The recommended system should be able to correctly pronounce the word with accurate ordinals such as first, second, third, fourth and distinguish between a year and a normal numeric value. In addition, the system should be able to recognize proper roman numeral usage, such as John II and Lecture II.

2. Find the right pronunciation in the content for various scenarios.

   There are some words in the text that have a different audible sound but the same characters. For example, the present tense and past participle of the word "read" can be pronounced differently. As a result, these issues must be addressed and a specific algorithm or a solution must be brought to address these kind of problems.

3. From written language, determine the proper accent, tension, and duration.

   The system should also be capable to identify various kinds of emotions, a tone which is addressed from the text, the feelings as the autistic kids find it difficult to differentiate the differences between these numerous emotions. The proposed system must be an assistance for those kind of ventures.

4. Implement a way for pronouncing the given text in a rhythmic and melodic manner.

   Representing a given context in a more meaningful way is another major factor that was paid attention to. The given content should be pronounced in the proper rhythm, with punctuation characters such as commas and question marks correctly identified. If the given text is a question, it should be spoken as such.

5. Availability for the growth

   The rational dealers who are deployed through the suggested platform can be functional and effectively engage with their defined objective 24 hours a day, 365 days a year. Because people are emotional beings, their cutting-edge intellectual renown can have an immediate impact on the quality of service provided by humans. The reasonable agent, on the other hand, believes that this incident will never happen.

6. With the proposed technology, you can manage a large number of requests at once.

   Smart marketers might be deployed using the proposed platform based on the current volume of requests to be handled, and shops could be able to handle many conversations at once, compared to their human counterparts.

7. Evaluate the evidence that give dealers have a more trust.

   In sales, the value of trust cannot be overstated. You may know everything there is to know about your product and have a lot of experience in this field, but if potential purchasers don't think you're trustworthy, you'll never be successful. You can provide all the discounts in the world on your incredible product features, but if you don't build confidence, you won't be able to meet your commercial or commercial objectives.

8. The service's average output is surging.

   When smart sellers are deployed through the proposed platform, they will be reactive and environmentally friendly. Customers of the machine can benefit in

productivity, time, and scalability due to the smart agent's capacity to interact with multiple customers at once.

Despite the wide range of speech synthesis technologies available, they all have drawbacks. As a result, the goal of this study is to discuss current ways to synthesizing human speech and to thoroughly examine its shortcomings and drawbacks and envelope the most appropriate solution to assist the autistic kids in helping out to sort their physical impediments. As a starting point, we looked at over a hundred recent papers that contained the most important findings from research on voice synthesis systems design, development, and implementation. Using this information, we were able to diagnose weak points in speech synthesis systems independent of the system design strategy used, and we were able to identify concerns that needed to be addressed in the near future.

4. RESEARCH METHODOLOGY

4.1 Requirement gathering and analysis.

The most important component of every software development project is acquiring and analyzing requirements. Because the solution to the problem is unknown, obtaining and analyzing requirements is critical, especially when it comes to research. As a result, a need should be explicit, realistic, and correct. It is critical to have a thorough understanding of the study topic. It is critical to establish that the suggested system is the best solution for the identified problem before moving on to the design and implementation phases.

As a result, a thorough requirement collection and analysis procedure is required. The project's requirements are collected by consulting a physician and then we launched a study in order to understand the speech impediments of the ASD kids within the range of 1-4 years old. All of the requirements that have been gathered and analyzed must be stated and recorded in this step.

4.2 System Overview

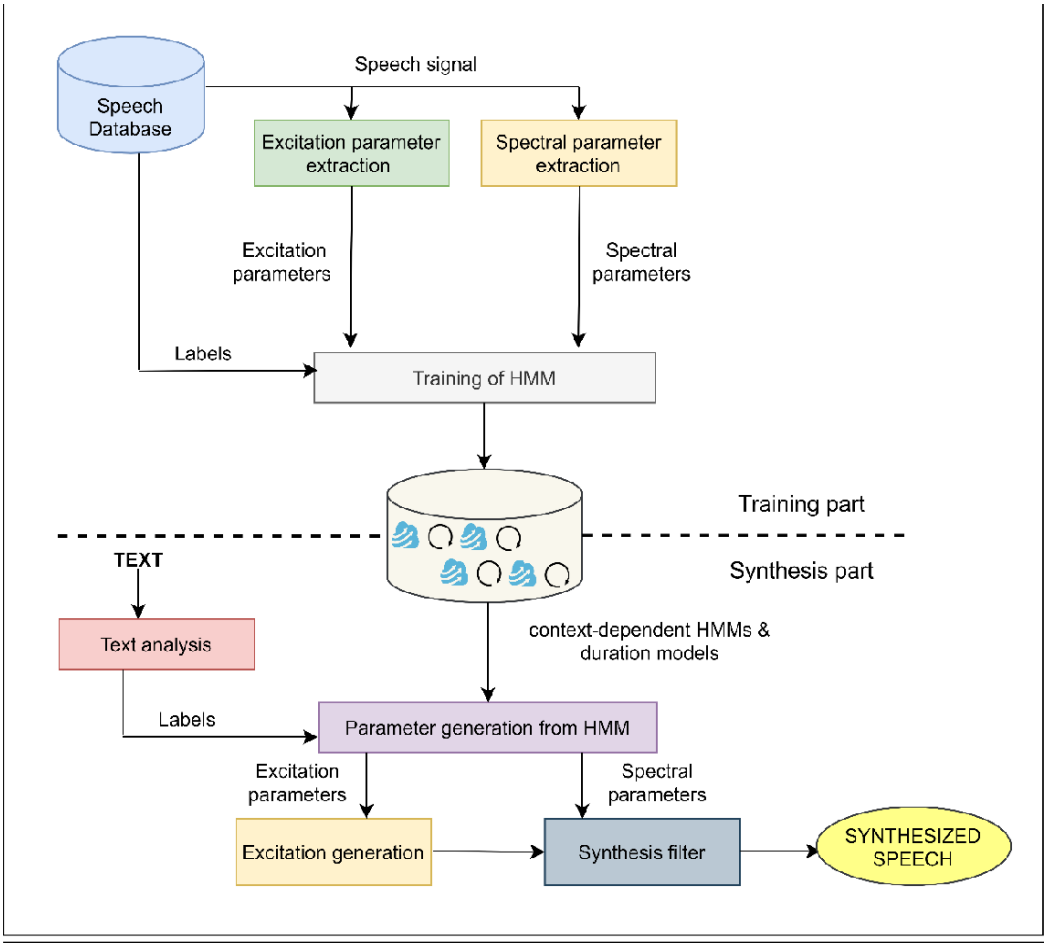


Figure 4. 1 Component Overview Diagram

4.3 System Process

Because of its tolerance to suboptimal recording circumstances, integrated data-driven technique for synthesizing units absents from the training data, and the possibility of speaker adaption, we chose HMM-based synthesis to produce voices. Using three different size subsets of the Child dataset as the target speaker corpus, we initially created two types of HMM-based voice synthesizers: speaker dependent and speaker adaptive. The suggested system is divided into two halves. The key portions are the training and synthesis sections. In the training module, there is a speech database that is utilized to retrieve excitation parameters and spectral parameters. The system will then be trained using HMM. Context-dependent HMMs and duration models are used

in the synthesis section. The given text will be processed first, and the variable will be created using HMM. The obtained stimulation parameters and spectral parameters will next be used to construct the synthetic speech.

In this section, many sorts of voice synthesis technologies for better text-to-speech conversion are discussed. The system is likely to acquire structure when both intelligibility and naturalness determine the maximal to generate artificial sounds of speech. The formant synthesis output speech is more likely to have a robotic and artificial sounding voice.

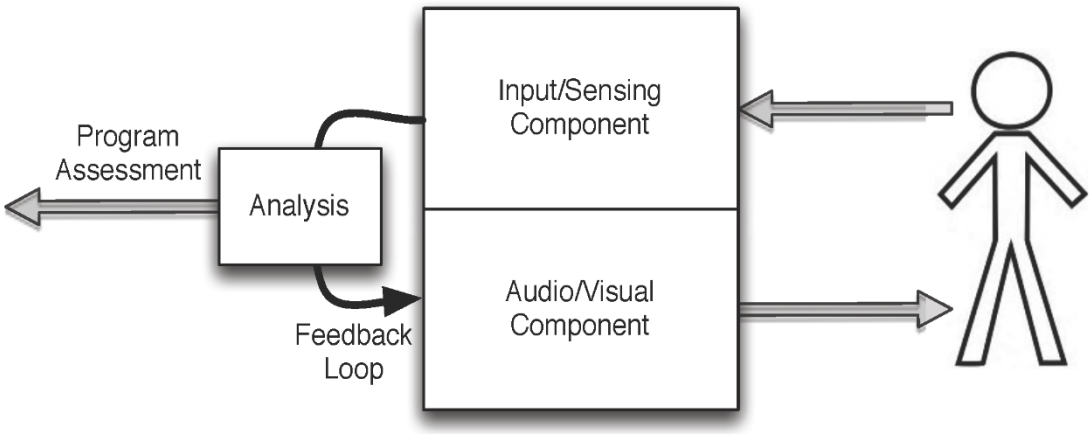


Figure 4. 2 Speech Process

The speech synthesis system based on HMM. It is divided into sections for training and synthesis. The EM algorithm is used in the training portion to perform maximum likelihood method. The fundamental distinction between this procedure and speech recognition is that both spectrums, mel-cepstral coefficients, and their dynamic characteristics and excitation parameters are taken from a database of speech patterns and modeled by a series of multi-streams context-dependent HMMs. In addition to phonetic contexts, linguistic and prosodic factors are taken into account.

The maximizing is done in the synthesis section. This might be thought of as the inverse of speech recognition. The utterances HMM is built by recombining the context dependent HMMs as per label sequence after converting a given sequence of words into a contextual label sequence. Second, the speech parameter synthesis algorithm takes the utterance HMM and generates spectral and excitation parameter patterns. Despite the fact that there are various varieties of the speech parameter generation algorithm,

the Case 1 algorithm is the most commonly utilized. Finally, utilizing excitation production and a speech synthesis filter, a voice waveform is synthesized using the generated spectral and excitation parameters. The voice parameter generating algorithm is detailed in the following sections.

Speech synthesis using a hidden Markov model (HMM) is a statistical parametric technique. It comprises of two phases: training and synthesis. Spectral and excitation features are retrieved, and HMMs are trained during the training phase. These HMMs can be context-agnostic or context-dependent. Provided a text, context independent or -dependent labels are produced during the formulation phase. Appropriate HMMs are chosen, and spectral and stimulation features are created, which are then used to synthesis speech. The mel-generalized cepstral coefficients, as well as their first and second counterparts, are used as spectral characteristics. These characteristics have a total of 105 dimensions. The log frequency response and its derivatives are the excitation properties. Five states are used to train the phoneme models, each with a single mixture component. In, you'll find a detailed overview of HMM-based speech synthesis. As noted in the introduction, an HMM-based synthesizer programmed with one hour of data can produce fairly decent synthesized speech. The current study only uses one hour of data to do the analysis for various phone sizes.

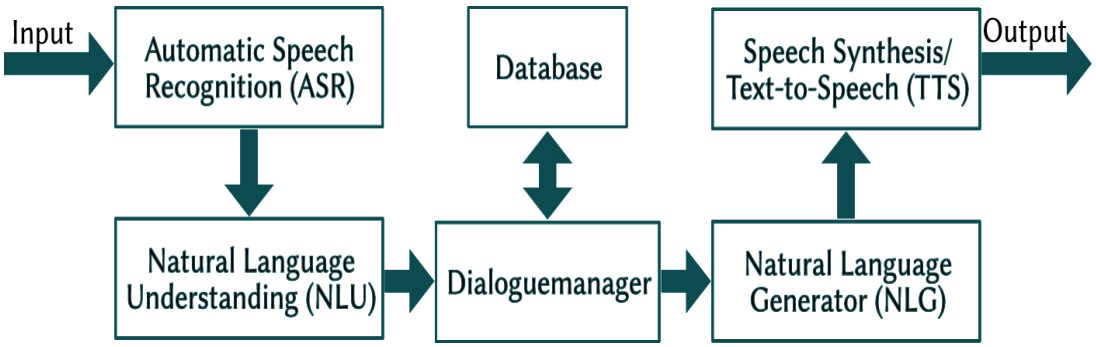


Figure 4. 3 Speech Generation from a random input

The above diagram depicts how a random text is taken as an input and how that particular voice is methodically generated to a speech .

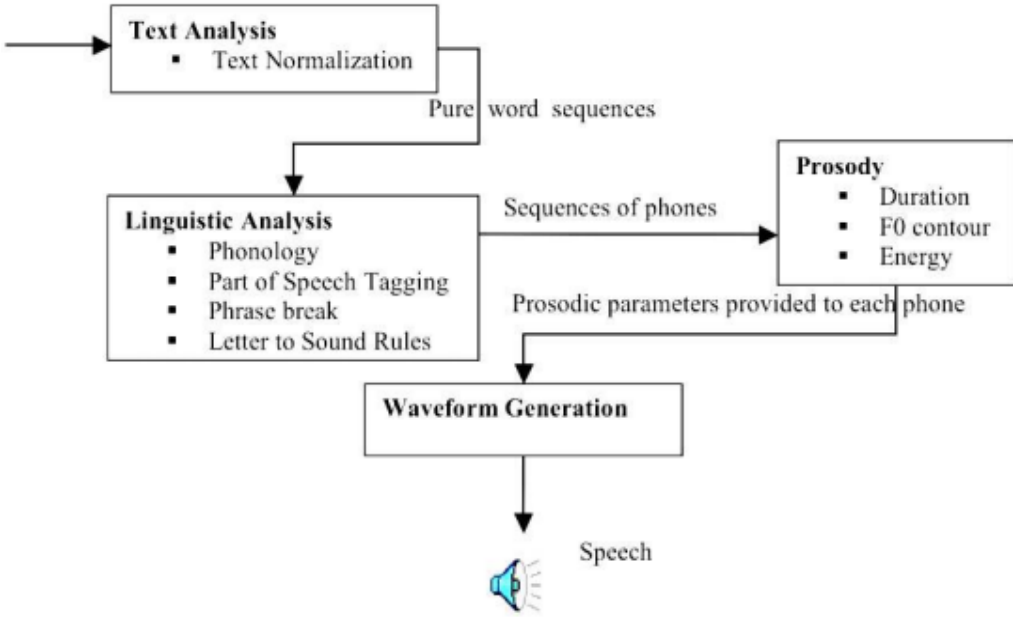


Figure 4. 4 Speech Generation Methodically

There are two types of Text-to-Speech systems. The first is limited domain TTS, while the second is generic TTS. TTS is a limited domain that was created with a specific purpose in mind. As a result, the application of this type of TTS is minimal. For voice synthesis, this TTS uses a limited number of words and sentences. A talking dictionary, a travel information system, a talking clock, and other similar applications are examples of this type of TTS. A general TTS, on the other hand, is designed to read anything from a document, including news from a website, emails, articles, and so on.

4.4 Phonotactics of Sinhala Language

Because the Sinhala TTS component of the project is so important, a detailed investigation of the Sinhala language and its properties that can affect the "ASD.AI" was done. The phonotactics of Sinhala speech can be displayed as below,

4.4.1 Diphthongs

The following Diphthongs are used in Sinhala. The secondary vowel in such diphthongs is always a high vowel. All vowels, except the center vowel /E/, occur as the initial vowel: /iu/, /eu/, /u/, /ou/, /au/, /ui/, /ei/, /i/, /oi/, and /ai/.

Phoneme sequences	Diphthong	Example
/ivu/ /iv/	/iu/	‘කිවුවා’‘කිව්වා’
/i:vu/ /i:v/	/i:u/	‘රජීව’
/evu/ /ev/	/eu/	‘පෙවුවා’‘පෙව්වා’
/e:vu/ /e:v/	/e:u/	‘සේව’
/ævu/ /æv/	/æu/	‘නැවුවා’ ‘නැව්වා’
/æ:vu/ /æ:v/	/æ:u/	‘බෑවුම’ ‘ගෑව්වා’
/ovu/ /ov/	/ou/	‘ඔවුන්’ ‘පොව්ව’
/avu/ /av/	/au/	‘කවුද’ ‘කව්පි’
/a:vu/ /a:v/	/a:u/	‘සාවුරුද්දක්’
/uyi/	/ui/	‘බකුයි’
/u:yi/	/u:i/	‘දූයි’
/oyi/	/oi/	‘පූසොයි’
/o:yi/	/o:i/	‘රෝයි’
/ayi/	/ai/	‘කයි’
/a:yi/	/a:i/	‘මායි’
/eyi/	/ei/	‘බැලෙයි’
/e:yi/	/e:i/	‘ගේයි’
/æyi/	/æi/	‘ඇයි’
/æ:yi/	/æ:i/	‘ඈයි’

4.4.2 Nasal Words in Sinhala

Sinhala's pre-nasalized voiced stops have characteristics of both voiceless stops and nasals. Pre-nasalized voiced consonants are shorter in duration than nasals phonologically. The below table shows the prenasalised consonants was used in building the TTS system which is used to support the ASD kids.

Table 4. 1 Personalized consonants in Sinhala Language

Prenasalized consonants in Sinhala Language						
	nasal	**obstruent**	**prenasalized consonant**	**Unicode**	**translit.**	**IPA**
velar	ඬ	ග	ඟ	0D9F	*ṅga*	[ⁿga]
retroflex	ණ	ඩ	ඬ	0DAC	*ṅḍa*	[ⁿɖa]
dental	න	ද	ඳ	0DB3	*ṅda*	[ⁿd̪a]
labial	ම	බ	ඹ	0DB9	*ṁba*	[ᵐba]

In addition, there is no unique symbol for /E/ in the Sinhala alphabet. The vowel /E/ does not appear at the start of a syllable except in the conjugational variations of verbs produced from the verbal stem /kErE/ - (to do). Despite the fact that the consonant sound/j/ is represented by the letter (in the Sinhala writing system), this is not considered a phenomenon under study in Sinhala.

4.5 Sinhala Prosodic Features

4.5.1 Stress

Stress is described as one of the key segmental features of speeches in the perspective of linguistic analysis. Single vowels and consonants, as well as complete syllables, are affected, regardless of position. When a syllable is emphasized in Spoken Sinhala, it is uttered with more force than an unstressed syllable. Also, the first syllable of a word is frequently more stressed.

4.5.2 Pitch

Pitch is another prosodic aspect to consider when using Sinhala TTS. It's the frequency at which the voice cords vibrate. The listener perceives differences in pitch as variations in vibration frequency. The higher the pitch in spoken Sinhala, the more frequently the vocal folds open and shut. This helps not only to announce variations in meaning, but also to recognize particular subclasses of words, mostly linked to verbs. Finite and nonfinite participles, as well as Imperative and Infinitive verbs, are the two types. The disparities in pitch that those subclasses are connected with are indicative of their distinctions.

Finite & nonfinite participle

: Example:

අමල් ගායනය කරයි (Amal is singing)

The finite participle which is a finite verb is pronounced with a falling intonation.

ගායනයෙන් පසු අමල් ගෙදර ගියේය (Amal went home after singing)

The nonfinite participle, which is a non-finite verb, uses a level intonation

4.5.3 Intonations

The term intonation is used to describe how pitch changes during the course of a phrase or sentence. The speaker's attitudes or feelings are conveyed through intonation. In other terms, intonation has a deictic purpose in discourse, such as inquiries, and a commonly understood function, such as wrath, sarcasm, or a variety of feelings. Intonation can also transmit simply grammatical information, such as indicating the end of a phrase. In Spoken Sinhala, three basic forms of intonation can be seen.

Falling intonation refers to a change in pitch from an upper to the lower pitch. The notion of finality is conveyed by expressions with falling Intonation.

අම්මා කිලිනොච්චියට ගියා - Mother has gone to kilinochchiya

Rising intonation: This signifies a change in pitch from a lower to a higher tone, conveying the idea of astonishment and unpredictability.

අම්මා කිලිනොච්චියට ගියාද? – Has mom gone to kilinochchiya

Level intonation: Level intonation illustrates the concept of non-finality by maintaining the same pitched ranges.

අම්මා කිලිනොච්චියේ උයනවා - Mom is cooking in kilinochchiya

4.6 How to Build a Synthetic Voice from a Text using the HMM?

Disfluencies, yelling, humming, sighing, paper turns, and other nonverbal noises were captured without interruptions, therefore the data had to be separated into shorter

segments. These elements were not attempted to be included into the synthetic voice. The information was transcribed by hand using standard orthography. Mispronunciations and word segments were dealt with great care so that the final phonological transcription accurately reflected the contents of the audio recordings. When a word in the vocabulary corresponded to the speaker's mispronunciation, it was utilized in the transcription. For example, the speaker frequently mispronounced the Sinhala phonetic word “ඇඹ” as "ඇඹ," hence the second word was used in the transcription. When there was no existing lexical item to match the speaker's articulation of a word or segment, the regular spelling transcription was replaced with a generated word, which was then entered to the vocabulary with the speaker's accent before the phonetic transcription was made.

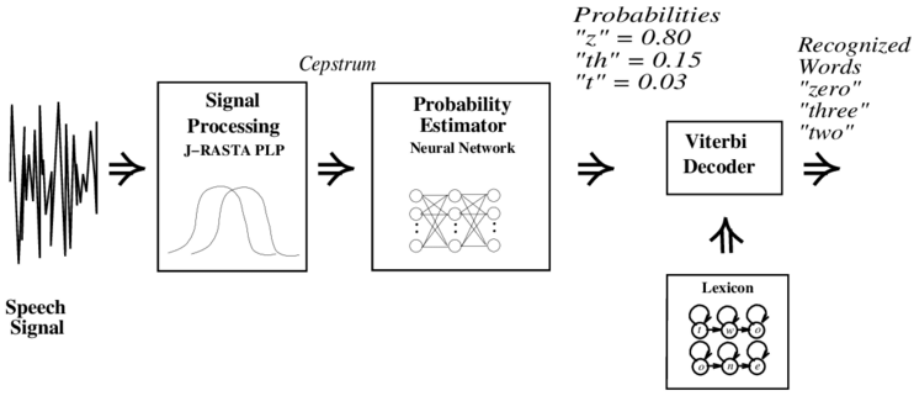


Figure 4. 6 Speech Signal Converted to recognizable word

The above diagram shows how a speech signal is converted to any sort of recognizable word with the aid of a vocabulary.

4.6.1 Training Stage of HMM

Figure below shows a schematic diagram of the HMM-based TTS system. The system is divided into two stages: training and synthesis.

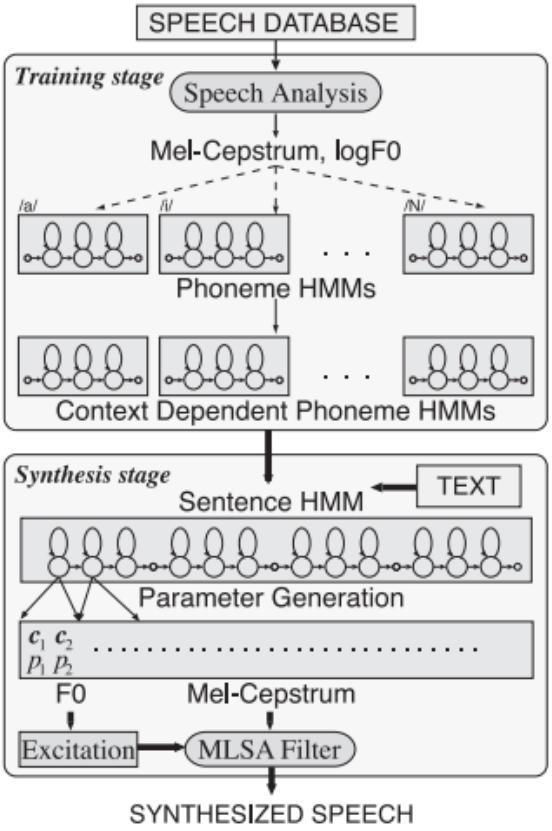


Figure 4. 7 Synthesis Speech System

Context-dependent phoneme HMMs are developed in the training step. Using a voice database is a good idea. At each evaluation, the spectrum and F0 are retrieved. Adaptive streaming HMMs are used to describe the output distributions for the spectral and logF0 sections, with a continuous random variable and a multi-space probability distribution (MSD) being used for the spectral and log F0 parts, respectively. We use the following phonological, prosodic, and linguistic data to describe fluctuations in the spectrum and F0. taking into account the following contexts:

- The number of morae in a phrase
- the location of the breath grouping in a paragraph; the number of morae in the breath groups preceding, current, and succeeding;
- The present accentual word's position within the current breath group;
- The number of morae in the previous, present, and subsequent dialectal phrases, as well as the type of emphasis;
- The preceding, current, and succeeding morphemes' parts of speech;

- The previous, present, and succeeding morphemes' parts of speech; of emphasis phonemes previous, present, and following;
- Fashion (for style-mixed modeling only).

4.6.2 Synthetic Stage of HMM

First, a randomly given text is turned into a series of context-dependent phoneme labels in the synthesis stage. A phrase HMM is built by concatenating context-dependent phoneme HMMs based on the label sequence. The ML criteria in which phoneme durations are computed using state time distributions, is used to extract spectral and F0 parameter patterns from the phrase HMM. Finally, voice is synthesized from the produced mel-cepstral and F0 parameter sequences using an MLSA (Mel Log Spectral Approximation) filter.

4.6.3 Speaker Conversion

Speech synthesis systems should, in principle, be able to reconstruct speech with any speaker features and speaking patterns. For example, in voice translation systems that are utilized by multiple speakers at the same time, it is important to reproduce the features of the input speakers so that listeners can recognize the speakers of the translated speech. Other example is multi-agent spoken dialog systems. Each person should have his or her unique speaker attributes and speaking styles for such devices. A number of spectral/voice translation approaches have been presented from this perspective. Because speech parameters employed in the developing solution are statistically modeled using the framework of the HMM, we may simply adjust spectral and prosodic properties of synthetic speech by converting HMM parameters suitably in the HMM-based speech synthesis approach.

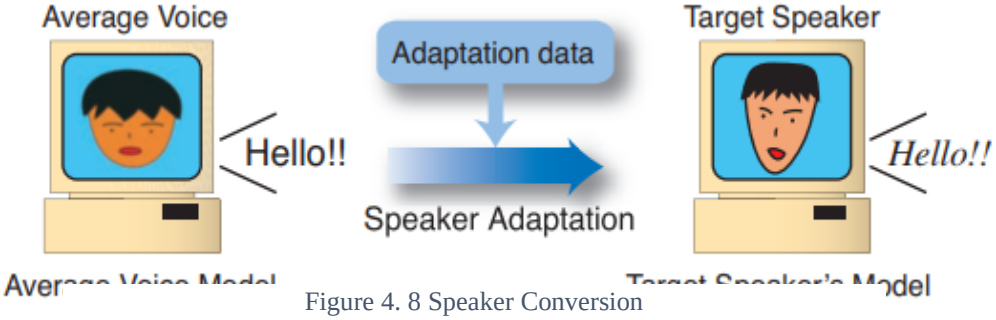


Figure 4. 8 Speaker Conversion

Basic models, such as mean vectors of output patterns, are modified to a target speaker using a limited amount of adaptation information uttered by the particular speaker in the speaker adjustment process. The original sample can be reliant or dependent on the speaker. Because most speaker adaptation techniques fail to work adequately between two participants with significant differences in voice features, it is necessary to select the speaker used for training the initial model adequately depending on the target speaker in the case of a speaker dependent initial model**.** Speaker adaptability approaches, on the other hand, function effectively for most target audiences when using speaker independent initial models, however the performance will be lower than when using speaker dependent beginning models that fit the target speaker and have enough data. We refer to the text - independent model as the "average voice model," and the synthesized speech created from the ordinary voice model as "average voice" because the synthetic speech produced from the speech processing model has averaged voice characteristics and speech sounds of speakers used for training..

4.7 Commercialization of the Product

The price of a speech-generating system might range from hundreds to thousands of dollars. A autistic kid should see a speech pathologist to determine the best device for them, program the device with their own words, and learn how to do it well. According on the expert offering the consultation, Medicare may reimburse up to 20 sessions of seeing a therapist about utilizing a speaker system. A portion of the consultation charge may be covered by some private health insurance plans. The governments worldwide have now been shifting into new ways of helping the differently abled kids with autism in order to support them. People with visual impairments or reading challenges can use text-to-speech software to listen to words written on a smartphone or computer. When a text-to-speech system is combined with a screen reader, a visually impaired user can use an auditory interface to interpret and conduct computer tasks. As a result, this system functions as an assistive device that enables these individuals to take advantage of information and communication technologies (ICT)

During the coming quarters, services are predicted to account for the majority of revenue. Text-to-speech software's functionality is dependent on services. They are an important part of the tool deployment process and are handled by solution, platform, and service providers. To deal with the constantly rising audio/video-based material, leading firms in numerous industries are using text-to-speech. This is assisting businesses in discovering new methods to tap into the vast amounts of data available in order to produce new products, services, and processes, so gaining a competitive advantage.

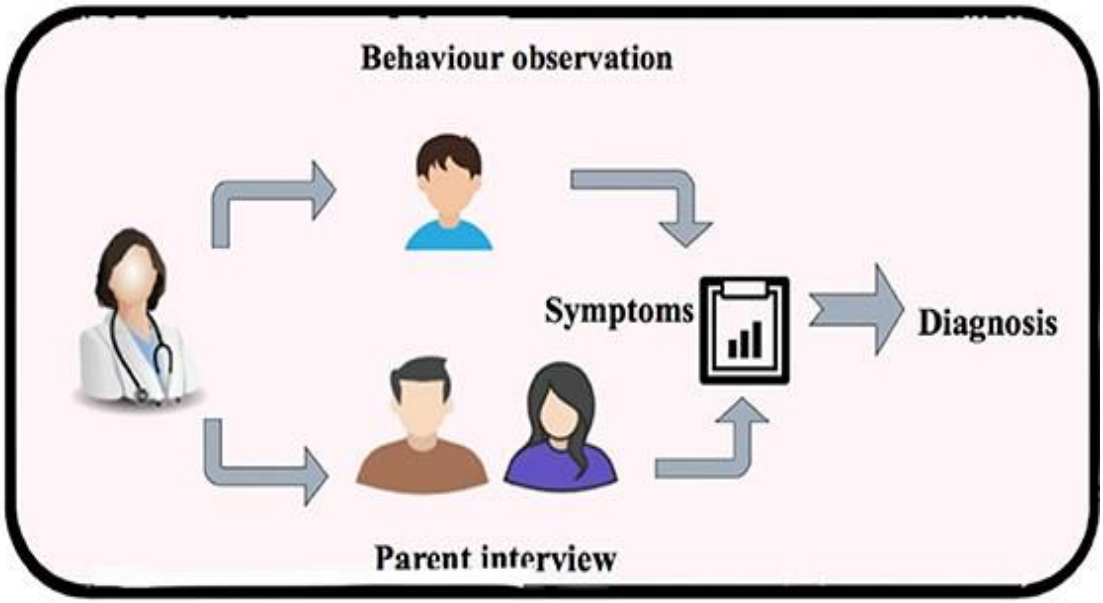


Figure 4. 9 ASD identification through consultation

In the process of creating a unique brand name for the commercialization process Voices developed in this way can be educated to produce a certain style of speaking in additional to sounding natural and fluent with proper rhythm and intonation. This comprises different syllable, phoneme, and word variants and inflections. Speech can be altered and changed using the Speech Synthesis Markup Language (SSML) by inserting items such as:

- Pauses
- Date
- Statistics

- Pronounce the acronym.
- Any more instructions on how to pronounce something.

We can also utilize these alternatives to establish a unique voice for our proposed solution , giving it a better and stronger presence in listeners' ears. We also may go even farther and train a custom model based on your own speech recordings to give your company a distinct sound. A brand's voice can convey a lot to listeners. As our brain is still developing an ear for sound, structure, and shape, we can extract specific information and meaning from the speaker's inflections or intonations when we attend. Some context-relevant characteristics, such as sarcasm, are considerably easier to transmit through audio than through writing.

For the system to reach a specific audience commercial advertising This strategy can be utilized to reach a larger audience at differing stages. Advertisements in newspapers are also successful and can be aimed to a wide range of characteristics. The following social media sites can be used to build a digital marketing platform.

- Facebook
- LinkedIn
- Instagram
- Twitter

Facebook is an online networking and digital social media platform based in the United States that connects people from all over the world by providing chat apps for global communication. Advertisements on Facebook can inform people about the "ASD.AI" system . On Facebook, the initiative and its goal can be easily shared. Furthermore, because Facebook allows for the creation of a business page, “ASD.AI” is interested in creating one in order to increase the number of users for the application. As a firm, we find that LinkedIn is the most effective way to market our product. We can also use ‘Google AdSense,' which is a Google-provided paid ad solution. This is a collection of styles for website authors, including text, video, and images, in a range of target kinds. Visitors can earn money by seeing or clicking on the advertisement.

4.7.1 Customer Relationships

The amount of consumer data that can be analyzed thanks to the practical use of the solution that we produced is enormous. As a result, organizations must first develop systems and methods for analyzing vast amounts of structured and unstructured data created by the device's operational use. The business purpose for evaluating consumer data, which in our instance is to generate marketing outcomes, is the most important component in creating the systems and processes. After the target has been determined, the analysis tools can be selected. As the application is specifically developed for the small kids from the age range of 1-4 yrs the marketing and commercialization aspect of the system should also be very tactical so that the parents of those kids would use our product over any other conventional devices existing in the market.

Apart from identifying the necessity to establish a n application for evaluating consumer data, it's also critical to determine the frequency with which the data should be examined, i.e., should it be done more regularly (daily, weekly, or monthly) or once a quarterly or every year? The goals of the company and the features of the product might determine the frequency of analysis.

This solution is primarily categorized as an industrial solution. As a result, revenue opportunities are limited.

Because the application is developed exclusively for autistic kids in Srilanka within the age range of 1-4 yrs , expected revenue streams include:

1. Contributions from organizations affiliated with Childcare and special organizations that assist children suffering from Autism.

2. The country's non-governmental organizations (NGOs).

3. After the application is introduced, purchase from both the domestic and international markets (Entire solution).

4.8 Tools and Technologies

- Python

Python is a high-level, general-purpose programming language that is translated. The use of considerable indentation in its design approach promotes code readability. Its language elements and object-oriented approach are aimed at assisting programmers in writing clear, logical code for both small and large-scale projects.

Figure 4. 10 Python Logo

- RASA Framework

Rasa is a platform for creating advanced manufacturing chatbots that are driven by AI. It's quite sophisticated, and developers all over the world use it to build chatbots and contextual assistants. We will learn some of the most crucial fundamental components of the Rasa framework and chatbot creation in this project. It's easy to construct conversational AI with Rasa and refine it over time. What's the best part? It uses an active learning method to help you improve, which is more convenient and spares you effort, cash, and resources. It can be used to quickly create excellent text- and voice-based chat bots. Rasa claims to provide a common infrastructure for AI chatbots, allowing all developers and businesses to quickly comprehend and construct their own virtual assistants.

Figure 4. 11 RASA Framework Logo

- Google Collab

Figure 4. 12 Google Collab Logo

Collaboratory, or 'Collab' for brief, helps in developing and execute Python code in your browsers with the following features: -

1. No configuration is necessary
2. Easy accessibility to GPUs
3. Simple sharing
4.

- Visual Studio Code

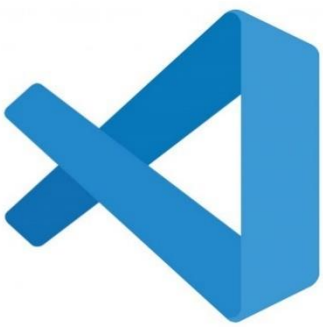


Figure 4. 13 Visual Studio Code Logo

Visual Studio Code is Microsoft's source code editing application, developed for Windows, Linux, and OS X. It is an application used an embedded Git control, debugging, syntax highlighting, excerpts, smart code completion, and code refactoring help. It us an agile, cross platform, and Multilanguage editor which can be downloaded directly from the website pf Visual Studio.

- **Android Studio (Includes Speech-to-Text library)**

Android Studio is Google's Android platform's integrated development environment. An y Apple, Windows and Linux operating systems are compatible with versions of Android Studio. With support for Google Cloud Platform and the integration of Google Apps, Android Studio offers developers a well-stocked toolkit to build Android apps or other projects and was an integral part of Android development since 2013. The speech to text library provides access to the speech recognition service which allows access to the speech recognizer. The class should not be called directly and instead, the Speech Recognizer create Speech Recognizer (Context) should be called. The methods of this class must be invoked only from the main application. The implementation of this API streams the audio to remote servers to perform the speech recognition.

Figure 4. 14 Android Studio Logo

5. Testing & Implementation

Following the development of synthetic speech, the quality of the speech is assessed using the metrics of clarity, naturalness, and applicability. Speech intelligibility with a high speech rate is frequently more significant than naturalness in some situations. Most assessment methods are designed to assess speech quality in general, but they are also applicable to synthetic speech. It's difficult, if not impossible, to say which test procedure produces the most accurate results. The ultimate speaking clarity in a text-to-speech system is determined not only by acoustic features, but also by text pre-processing and syntactic realization. The testing part of the system should start from creating the test cases stage until it reaches the system testing stage. This can be shown as below.

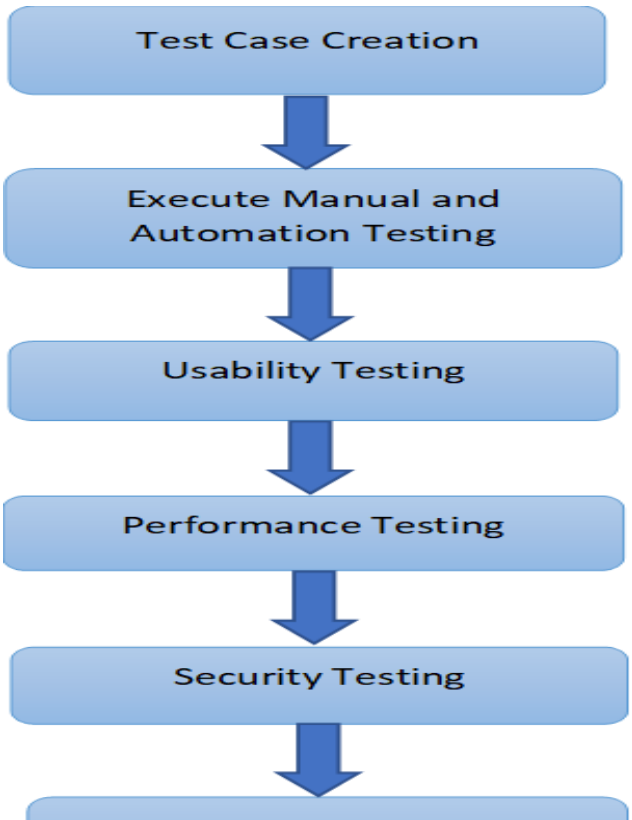


Figure 5.1 Testing Types

5.1 Types of Testing

1. Performance Testing

   The application's Interfaces are assessed for performance and user friendliness.

2. Device Testing

   The device is tested across many sample of kids and their interaction with the system is taken into consideration in order to launch the final output.

3. Maintained Test case Sheet

   Retained a test case sheet for each single component, combining all test case sheets for each aspect to provide a test case report for the “ASD.AI” application.

5.2 Implementation

- Identify the requirements of ASD.AI

  A meeting with a doctor who treats dyslexic kids from the age range of 1-4 yrs old who are suffering with the Autism spectrum disorder was an excellent opportunity to learn about the application's needs and then creating a proper plan to address the ongoing problems. Furthermore, the necessity for "ASD.AI" requirements is revealed by certain reading testing with ASD and non-ASD kids. A meeting with the professors was also held in order to have a better understanding on the subject area.

- Strategies

  Consider merging the software strategy with the business objective. Define an advantage, such as support for the native language (Sinhala), user empowerment, and so forth. The developed application is projected to provide a cost-effective and effective, user-friendly, and multi-sensory solution.

- Set the Scope

  Each team member created wireframes for the user interfaces and finalized the design of the user interfaces. User interfaces were developed, the entity relationship diagram was constructed, and the database was created. I began working on the back end of the project. Each team member put in a lot of attempt to complete the development.

- Implememtation Planning

  It was decided what kind of technologies will be used in the development. Developed the application based on the strategies and needs specified. The deployment is carried out in accordance with the time frame determined by the "ASD.AI" team.

5.2.1 Proof of Implementation

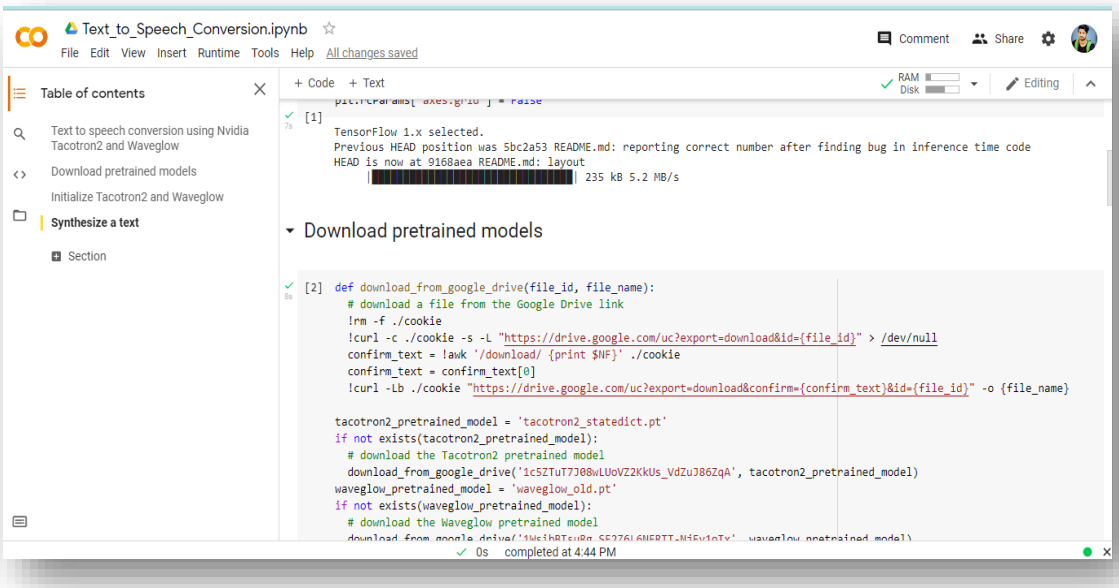


Figure 5.2. 1: Pretrained data for speech synthesis

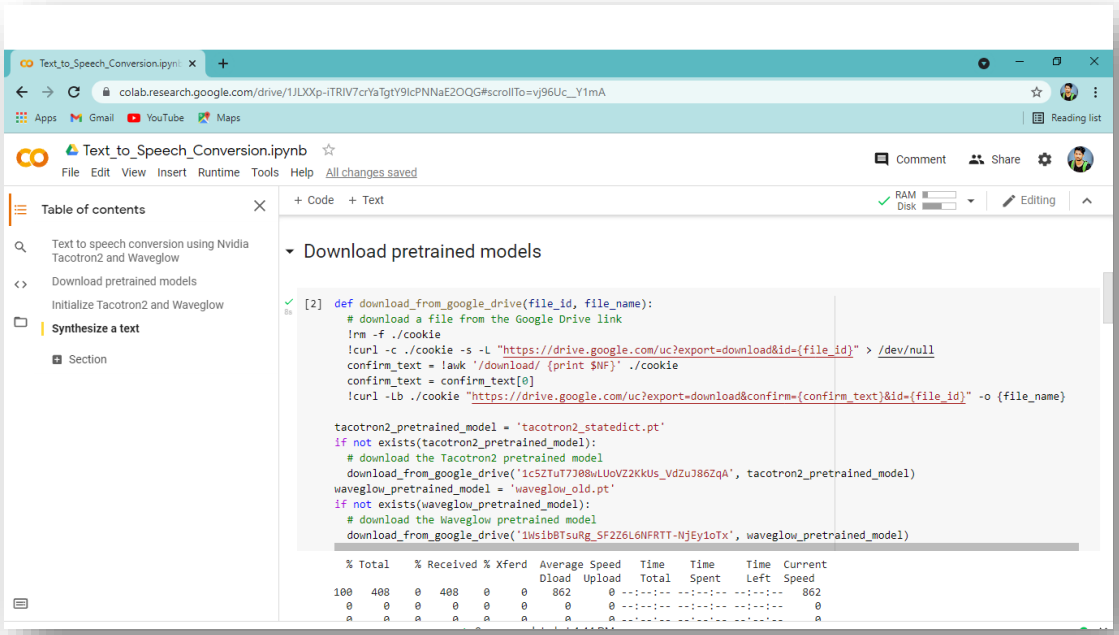


Figure 5.2. 2: Download the pretrained data model

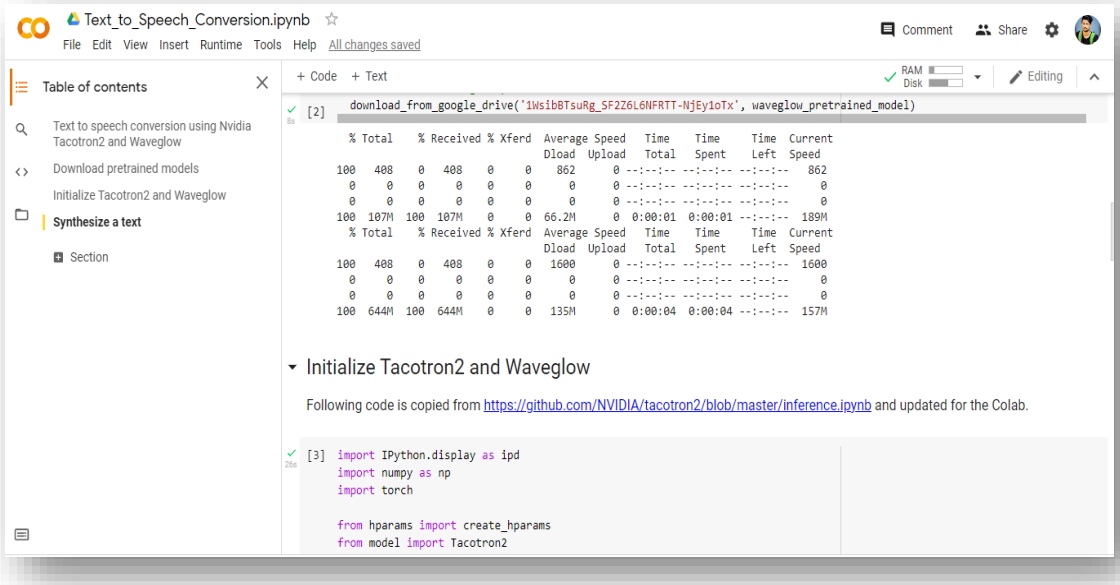


Figure 5.2. 3: Depiction of the downloaded pretrained data set

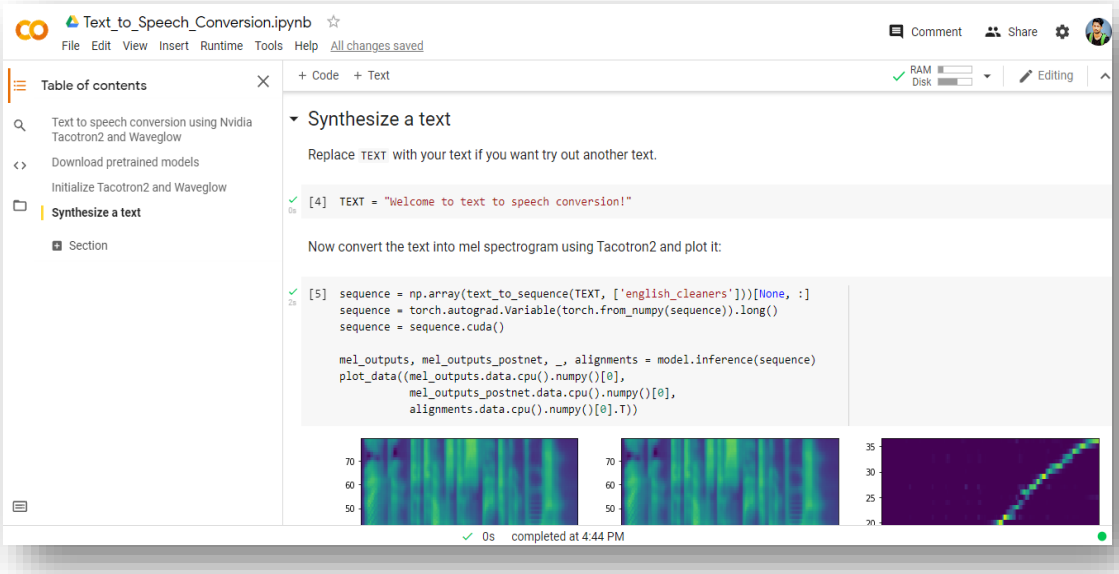


Figure 5.2. 4: Synthesized speech text depiction

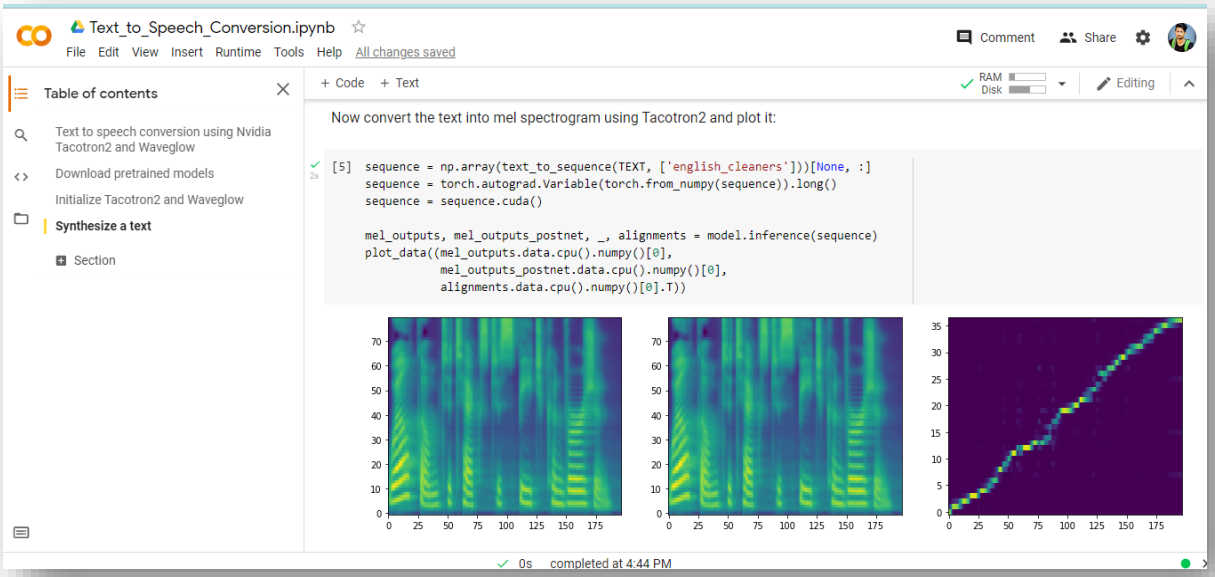


Figure 5.2. 5: Synthesizing a text illustration

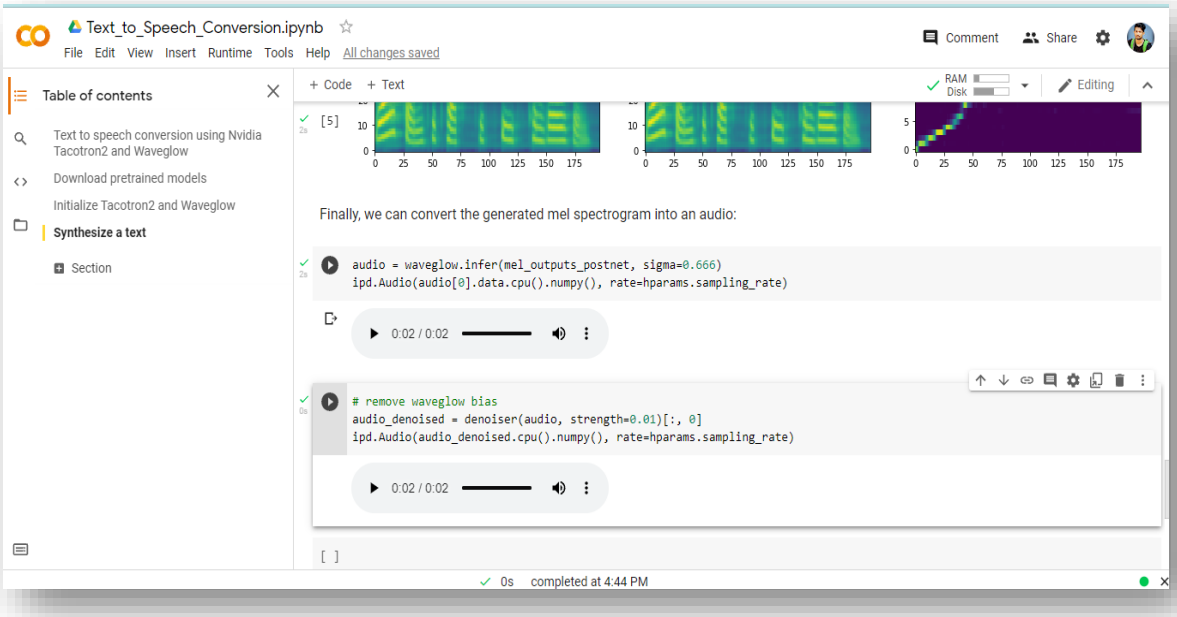
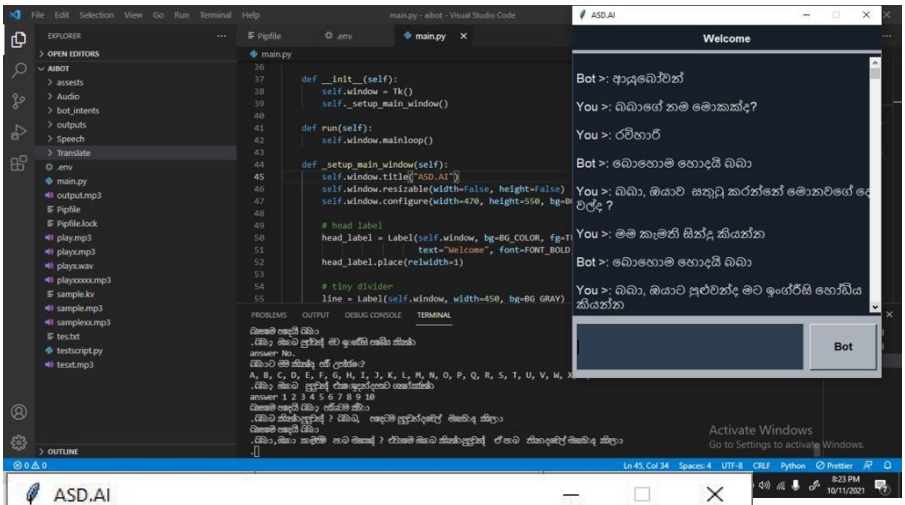
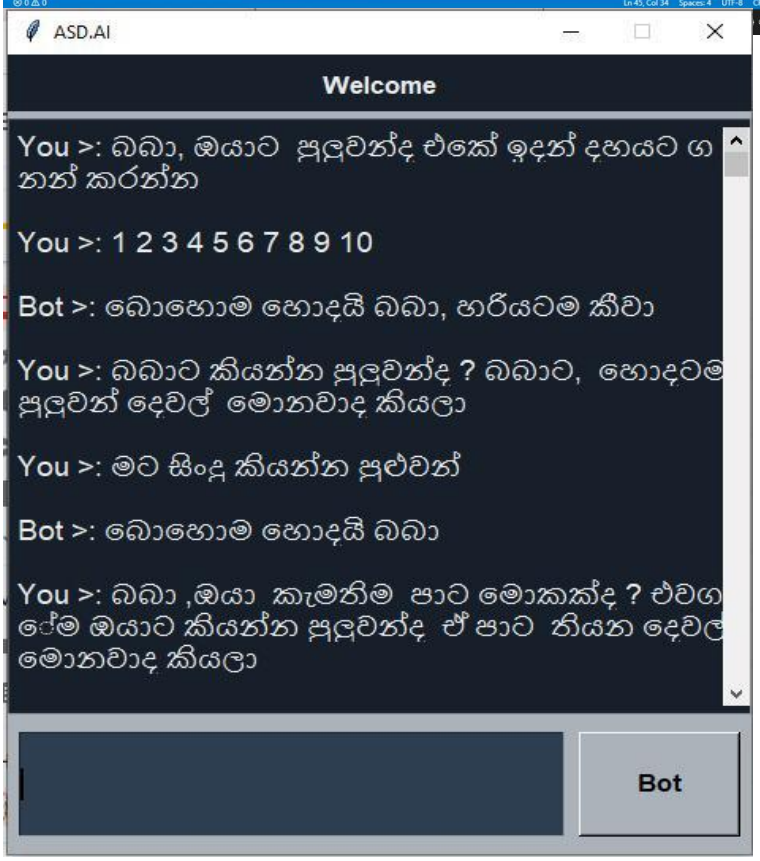


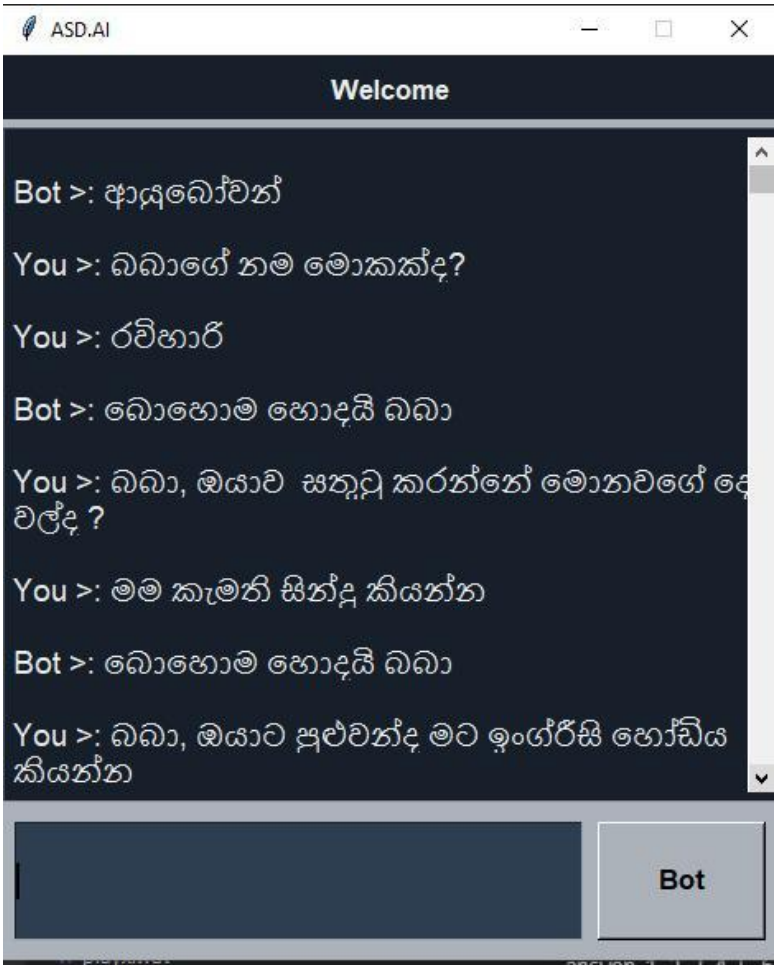
Figure 5.2. 6 :Text conversion to speech illustration

The above diagram shows the sinhala speech input and its output function in the window.



and the




ASD.AI
Welcome
Bot >: ආයුබෝවන්
You >: බබාගේ නම මොකක්ද?
You >: රවිහාරි
Bot >: බොහොම හොඳයි බබා
You >: බබා, ඔයාව සතුටු කරන්නේ මොනවගේ දෙ
වල්ද ?
You >: මම කැමති සින්දු කියන්න
Bot >: බොහොම හොඳයි බබා
You >: බබා, ඔයාට පුළුවන්ද මට ඉංග්‍රීසි හෝඩිය
කියන්න
Bot

5.3 Test Case Scenarios

5.3.1Test Scenarios for Login function

Table 5.3. 1 Test Case for Login Scenario

Test Scenario ID	Login -1	**Test Case ID**	Login 1A
Test Case Description	Login – successful	**Test Priority**	High
Pre-requisite	Should be valid User	**Post-requisite**	NA

Test Execution Steps:					
Step number	**Action**	**Inputs**	**Expected result**	**Actual result**	**Test results**
1	Launch the application	-	Login page is shown	Login page is shown	Pass
2	Click the login button	-	Login form is shown	Login form is shown	Pass
3	Enter valid username and password and click login button	Username: Admin Password: admin	Successfully logged in success message appears and redirected to the application	Successfully logged in toast message appears and redirected to the application	Pass

Test Scenario ID	Login -2	**Test Case ID**	Login 2A
Test Case Description	Login - negative	**Test Priority**	High

Pre-requisite	A valid password should be entered	**Post-requisite**	NA

Test Execution Steps:

Step number	**Action**	**Inputs**	**Expected result**	**Actual result**	**Test results**
1	Launch the device	-	Login page is appeared	Login page is appeared	Pass
2	Click the login button	-	Login form is appeared	Login form is appeared	Pass
3	Enter invalid username and valid password and click login button	Username: abcd3456Password: 7896%%%$$$	Invalid login message appears and ask the user is asked to try again	Invalid login message appears and ask the user to try again	Pass

Table 5.3. 2 Test case for Unsuccessful login Admin Side

Test Scenario ID	Login -2	**Test Case ID**	Login 2B
Test Case Description	Login - negative	**Test Priority**	High
Pre-requisite	Valid Username should be there	**Post-requisite**	NA

Test Execution Steps:

Step number	**Action**	**Inputs**	**Expected result**	**Actual result**	**Test results**

1	Launch the application	-	Login page is appeared	Login page is appeared	Pass
2	Click the login button	-	Login form is appeared	Login form is appeared	Pass
3	Enter valid username and invalid password and click login button	Username: Admin Password: &&&6789	Invalid login message appears and ask the user to try again	Invalid login message appears and ask the user to try again	Pass

Table 5.3. 3 Test Case for unsuccessful login- User side

Test Scenario ID	Login -2	**Test Case ID**	Login 2B
Test Case Description	Login - negative	**Test Priority**	High
Pre-requisite	-	**Post-requisite**	NA

Test Execution Steps:

Step number	**Action**	**Inputs**	**Expected result**	**Actual result**	**Test results**
1	Launch the device	-	Login page is displayed	Login page is displayed	Pass
2	Click the login button	-	Login form is displayed	Login form is displayed	Pass

3	Enter invalid username and invalid password and click login button	Username: &&& Password: &&&****	Invalid login message appears and ask the user to try again	Invalid login message appears and ask the user to try again	Pass

5.3.2 Test Scenarios for Speech Signal Processing function

Table 5.3. 4 Test Case for speech signal successful input

Test Case Id	Description	Inputs	Expected Outcome	Actual Outcome	Result
01	Enter the speech signal	User enters the specific speech text	The text should be read	Same as expected outcome	Pass

Table 5.3. 5 Test case for speech signal unsuccessful input

Test Case Id	Description	Inputs	Expected Outcome	Actual Outcome	Result
01	Enter the speech signal	User enters the specific speech text	The text is not read as expected read	Same as expected outcome	Fail

5.4 Types of Testing

The solution was put through functional and non-functional testing to assure its functional reliability. The technique by which QAs verify if a software program is behaving in accordance with the pre requirements is known as functional testing.

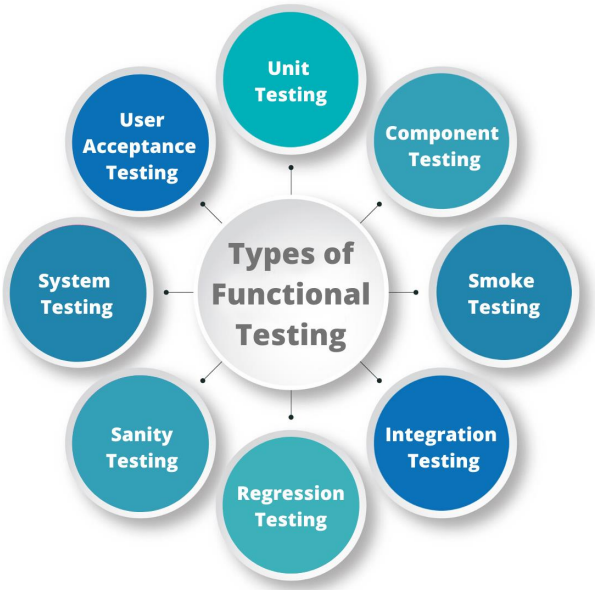


Figure 5.4. 1 Functional Testing

Table 5.4. 1 Functional Testing

Integration testing	To guarantee that each individual members integrated components were functioning correctly, an integration test was undertaken.
Unit testing	Individual components have been subjected to unit testing. Each component was put through its paces on its own. Individual components have been subjected to unit testing. Each component was put through its paces without regard to the others.. In the unit testing process the following steps should be followed; • Collect a variety of audio files.

	• Convert them into Conversation Time Mark (CTM) format using Speech to Text and combine them into one single file. • A list of all inputs and outputs, as well as an explanation of any problems and summarized statistic for the word error rate (WER). • A summary statistic called the sentence error rate (SER).

Non-functional testing is a sort of testing process that examines a software application's non-functional elements (performance, usability, dependability, and so on). It's intended to assess a system's readiness based on nonfunctional criteria that aren't covered by functional testing.

Figure 5.4. 2 Nonfunctional Testing

Table 5.4. 2 Nonfunctional Testing

Usability testing	Usability testing was conducted to guarantee that the system is useful and pleasurable to use for the autistic kids within the range of 1-4 yrs.
Performance testing	The frequency at which data and information are circulated through the devices and dashboard was measured in a performance test.
Security testing	• Scanning for active anti-virus and malware at the system level • Test for penetration • Passive: Watches network activity but doesn't generate any of its own. • Alerts
Compatibility testing	A compatibility test was performed to ensure that the input text readings was compatible with the output speech text.

6. RESULTS & DISCUSSION

6.1 Results

Speech synthesis algorithms can be assessed based on a variety of criteria, including speech perception, spontaneity, computational complexity, and so on . It's fair to presume that new assessment metrics will be required for acoustic intelligence applications, such as sentimental influence on the user, capacity to get the user to act, mastery of language generation, and whether the system considers environmental factors and adjusts its behaviors accordingly. TTS (text-to-speech) is a common assistive technology in which a computers or tablet reads out vocally to the user the letters and words. This technique is popular among kids who have literacy issues, particularly those who have trouble decoding. By providing the words in an auditory format, the kids may concentrate on the meaning of the phrase rather than using all of their mental resources to sound them out. While this technology can help students overcome their reading issues and gain access to school materials, it does not aid in the development of reading abilities. However, the research demonstrated that kids with ASD benefited from using the ASD.AI software. For six weeks, this team provided small-group software training to kids, and they witnessed gains in motivation to study, understanding, and pronunciation . Another study revealed that ASD.AI was helpful in assisting children to access reading content and was also well-liked by the students who used it, particularly from the ages ranging from 1-4 yrs old.

The assessment showed that HMM-based systems were chosen over voice conversion unit selection methods as being handier to the original speaker. However, there is a difference in interpretation. The fact that these biases for HMM-based systems over voice-based systems are complicated by the fact that

6.2 Research findings and Discussion

Ten differently abled children suffering with speech impediments children with ASD (5 females, 5 males), and 12 TD children (4 females, 8 males), age ranging from 4 to 6 years, as well as 5 TD younger children, aged 2 to 3 years (3 males), participated in the study. . All of the 4- to 5-year-old children with ASD in the final study had a verbal age of 40 months or more. According to parent assessments, all of the students spoke

Sinhala fluently and had little exposure to other languages. Data from two more participants who were originally enrolled for the ASD group were eliminated because their cognitive age was less than 40 months. Finally, because to extended exposure to a second language, data from one extra participant in the ASD group and two extra children originally selected for the TD group were eliminated. Using a HMM paradigm, we evaluated brain responses to speech and non-speech sounds in children with ASD, who were independently matched on verbal age. The purpose of this study was to learn more about the brain mechanisms that underpin speech recognition and processing in this population. This is the only ERP study that we are aware of that looks at the detection and discriminating of speech from non-speech in 4- to 6-year-old children with ASD without the use of oddball stimuli or accompanying attentional orienting responses.

6.3 Conclusions

Speech applications require testing issues that are distinct from those faced by other apps. It is possible and necessary to test the speech portions of your application separately from the rest of the application. This dissertation described how to evaluate the performance of both voice input (speech to text) and voice output (speech to speech) (text to speech). In conclusion it can be stated that the speech synthetic systems that we tried to implement in order to assist the ASD kids with the speech impediments will assist them in their childhood and help them to socialize with the society much better than they are doing now. Even seemingly minor changes in the quantity and similarity of extra data can have considerable influence on model performance in our scenario of data scarcity and domain specialization. Importance of data quality rather than quantity, with a restricted age range that matched the goal data outperforming other combinations with more variance and quantity. It was also founded that even with bigger amounts of data, model performance may be improved if domain similarity can be recognized and expanded upon. Our HMM based speech synthesis system will be tested in the future for a variety of additional health-related tasks, such as speech clarification identification. We'll also see how well the TTS system handles other typical behavioral inputs like video descriptors and physiological properties like ECG and electrodermal activity representations. We'll also compare and combine HMM based feature learning with other unsupervised representation learning approaches.

7. References

[1] H. Hodges, C. Fealko, and N. Soares, “Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation,” *Transl. Pediatr.*, vol. 9, no. Suppl 1, pp. S55–S65, 2020.

[2] F. Yaylaci and S. Miral, “A comparison of DSM-IV-TR and DSM-5 diagnostic classifications in the clinical diagnosis of autistic spectrum disorder,” *J. Autism Dev. Disord.*, vol. 47, no. 1, pp. 101–109, 2017.

[3] V. N. Vahia, “Diagnostic and statistical manual of mental disorders 5: A quick glance,” *Indian J. Psychiatry*, vol. 55, no. 3, pp. 220–223, 2013.

[4] M. J. Maenner *et al.*, “Potential impact of DSM-5 criteria on autism spectrum disorder prevalence estimates,” *JAMA Psychiatry*, vol. 71, no. 3, pp. 292–300, 2014.

[5] A. H. Memari *et al.*, “Children with autism spectrum disorder and patterns of participation in daily physical and play activities,” *Neurol. Res. Int.*, vol. 2015, p. 531906, 2015.

[6] A. Parmeggiani, A. Corinaldesi, and A. Posar, “Early features of autism spectrum disorder: a cross-sectional study,” *Ital. J. Pediatr.*, vol. 45, no. 1, p. 144, 2019.

[7] A. V. Kirby, M. L. Diener, D. E. Adkins, and C. Wright, “Transition preparation activities among families of youth on the autism spectrum: Preliminary study using repeated assessments across a school year,” *PLoS One*, vol. 15, no. 4, p. e0231551, 2020.

[8] L. Carroll *et al.*, “Autism spectrum disorders: Multiple routes to, and multiple consequences of, abnormal synaptic function and connectivity,” *Neuroscientist*, vol. 27, no. 1, pp. 10–29, 2021.

[9] S. C. Suzuki *et al.*, “Cadherin-8 is required for the first relay synapses to receive functional inputs from primary sensory afferents for cold sensation,” *J. Neurosci.*, vol. 27, no. 13, pp. 3466–3476, 2007.

[10] K. N. Thakkar et al., “Response monitoring, repetitive behaviour and anterior cingulate abnormalities in autism spectrum disorders (ASD),” Brain, vol. 131, no. Pt 9, pp. 2464–2478, 2008.

[11] S. C. Suzuki *et al.*, “Cadherin-8 is required for the first relay synapses to receive functional inputs from primary sensory afferents for cold sensation,” *J. Neurosci.*, vol. 27, no. 13, pp. 3466–3476, 2007.

[12] J. M. S. Gavaldá and T. Qinyi, “Improving the process of inclusive education in children with ASD in mainstream schools,” *Procedia Soc. Behav. Sci.*, vol. 46, pp. 4072–4076, 2012.

[13] L. Campisi, N. Imran, A. Nazeer, N. Skokauskas, and M. W. Azeem, “Autism spectrum disorder,” *Br. Med. Bull.*, vol. 127, no. 1, pp. 91–100, 2018.

[14] A. H. Memari *et al.*, “Children with autism spectrum disorder and patterns of participation in daily physical and play activities,” *Neurol. Res. Int.*, vol. 2015, p. 531906, 2015.

1[5] A. V. Kirby, M. L. Diener, D. E. Adkins, and C. Wright, “Transition preparation activities among families of youth on the autism spectrum: Preliminary study using repeated assessments across a school year,” *PLoS One*, vol. 15, no. 4, p. e0231551, 2020.

[16] L. Carroll *et al.*, “Autism spectrum disorders: Multiple routes to, and multiple consequences of, abnormal synaptic function and connectivity,” *Neuroscientist*, vol. 27, no. 1, pp. 10–29, 2021.

[17] M. Bulut and S. S. Narayanan, “Speech synthesis systems in ambient intelligence environments,” in *Human-Centric Interfaces for Ambient Intelligence*, Elsevier, 2010, pp. 255–277.

[18] R. Aida-Zade, C. Ardil, and A. M. Sharifova, “The main principles of text-to-speech synthesis system,” vol. 37, 2010.

[19] T. J. Sefara, T. B. Mokgonyane, M. J. Manamela, and T. I. Modipa, “HMM-based speech synthesis system incorporated with language identification for

low-resourced languages," in *2019 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*, 2019, pp. 1–6.

[20] N. P. Narendra, K. S. Rao, K. Ghosh, R. R. Vempada, and S. Maity, "Development of syllable-based text to speech synthesis system in Bengali," Int. J. Speech Technol., vol. 14, no. 3, pp. 167–181, 2011.

[21] *Springer.com*. [Online]. Available: https://link.springer.com/article/10.1007%252Fs11528-014-0825-7. [Accessed: 02-Oct-2021].

[22] S. L. Ness *et al.*, "JAKE® multimodal data capture system: Insights from an observational study of autism spectrum disorder," *Front. Neurosci.*, vol. 11, 2017.

[23] T. J. Sefara, T. B. Mokgonyane, M. J. Manamela, and T. I. Modipa, "HMM-based speech synthesis system incorporated with language identification for low-resourced languages," in 2019 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), 2019, pp. 1–6.

[24] M. Wester, Z. Wu, and J. Yamagishi, "Human vs machine spoofing detection on wideband and narrowband data," in *Interspeech 2015*, 2015.

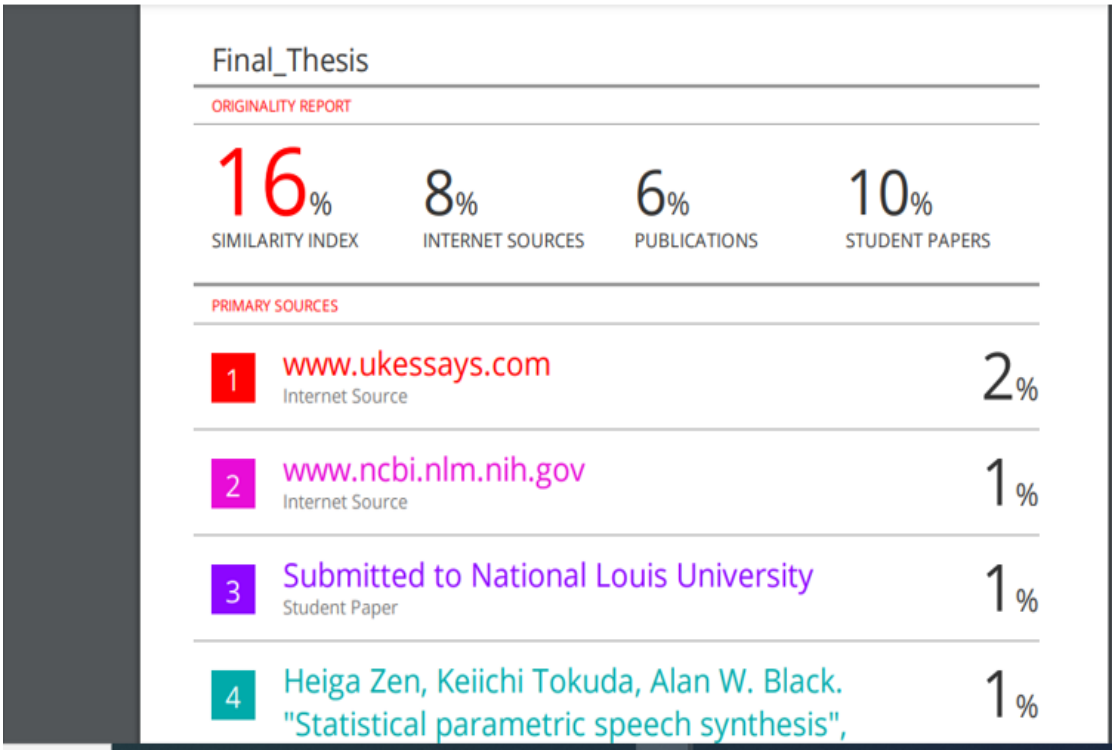
[25] *Autismspeaks.org*. [Online]. Available: https://www.autismspeaks.org/expert-opinion/autism-and-speech-devices-helping-kids-advance-skills-they-mature. [Accessed: 02-Oct-2021].

[26] N. Zeliadt, "New technology gives voice to nonverbal people with autism," *Spectrumnews.org*, 22-Jun-2015. [Online]. Available: https://www.spectrumnews.org/news/new-technology-gives-voice-to-nonverbal-people-with-autism/. [Accessed: 02-Oct-2021].

[27] F. Chen, L. Wang, G. Peng, N. Yan, and X. Pan, "Development and evaluation of a 3-D virtual pronunciation tutor for children with autism spectrum disorders," *PLoS One*, vol. 14, no. 1, p. e0210858, 2019.

Appendices

Appendix A- Plagiarism Content



Final_Thesis

ORIGINALITY REPORT

16%	8%	6%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ukessays.com Internet Source	2%
2	www.ncbi.nlm.nih.gov Internet Source	1%
3	Submitted to National Louis University Student Paper	1%
4	Heiga Zen, Keiichi Tokuda, Alan W. Black. "Statistical parametric speech synthesis",	1%