ASD.AI – SINHALA DIALOGUE MANAGEMENT TOOL TO SCREEN KIDS WITH AUTISM SPECTRUM DISORDER

Gunawardhana M.D.R.T.	IT16090804
Herath H.M.D.N.	IT18081794
Anjali R.P.D.N.	IT17109536
Sampath G.A.D.M.	IT16061880

Bachelor of Science (Honors) in Information Technology

Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

October 2021

ASD.AI – SINHALA DIALOGUE MANAGEMENT TOOL TO SCREEN KIDS WITH AUTISM SPECTRUM DISORDER

Gunawardhana M.D.R.T.	IT16090804
Herath H.M.D.N.	IT18081794
Anjali R.P.D.N.	IT17109536
Sampath G.A.D.M.	IT16061880

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

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

October 2021

DECLARATION

I declare that this is my own work and this dissertation I does not incorporate without acknowledgment any material previously submitted for a degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology, the non-exclusive right to reproduce and distribute our dissertation, in whole or in part in print, electronic or another medium. I retain the right to use this content in whole or part in future works (such as articles or books).

	13th Oct 2021
---	------------------
Signature (Gunawardhana M.D.R.T.)	Date

The above candidate has carried out research for the bachelor's degree dissertation under my supervision.

-----------------------------------	--------------------------
Signature of the Supervisor	Date
-----------------------------------	---------------------------
Signature of the Co-Supervisor	Date

ABSTRACT

Computer-based conversational systems are quite common in today's society. It might be a human-powered conversational system or a human-machine conversational system. Human-machine communication systems such as voice assistants are well-known. Some of the most popular and advanced voice assistants are capable of acting almost naturally. The bulk of voice assistants, on the other hand, are exclusively accessible in English. However, because to a language barrier, those who are unable to fully utilize the available voice assistants may lose out on obtaining support from conversational agents/voice assistants. Furthermore, users are not forced to have a formal conversation with a bot; alternatively, users may avoid conversing freely due to a lack of trust in asking questions.

This research seeks to create a solution for a specific domain, such as a Sinhala voice assistant. A generic voice assistant will almost certainly fail to respond to the majority of questions, leaving it worthless and ineffective. As a consequence, users may utilize a domain-specific voice assistant to ask questions and solve problems without wasting time pointing to a specific location. Building 'A Sinhala voice assistant for assessing youngsters with autism spectrum disorder' largely focuses on addressing two problems by being language specific (Sinhala) and domain specific.

Rather of answering simple questions, this voice assistant concentrates on identifying the objective of the person who has asked the inquiry and correctly responding. As a consequence, this voice assistant can handle context-sensitive conversations. The responses become more trustworthy and acceptable when compared to a conventional rule-based, template-based, or keyword-based bot. The voice assistant was built using the RASA platform, an open-source conversation management system. The framework's two main components are RASA Core and the RASA NLU, with RASA Core enabling users with more sophisticated discussions and chats. It enables interactive learning and supervised machine learning to be used to train the dataset. RASA NLU performs natural language processing tasks such as purpose classification and entity extraction in voice assistants/dialogue models.

ACKNOWLEDGMENT

Apart from the efforts of the project team, the success of any project rests largely on the concerted endeavors of many others. First and foremost, we would like to thank our supervisor Prof Koliya Pulasinghe for providing insight and expertise that greatly assisted the research and also for supervising our project. We also thank our co supervisor Ms. Vijani Piyawardana for her sincere and selfless support, prompt and useful advice during our research. We also thank our colleagues from Sri Lanka Institute of Information Technology who have given their invaluable support and comments for our research project to make it a success.

TABLE OF CONTENTS

DECLARATION 1

ABSTRACT 2

ACKNOWLEDGMENT 3

TABLE OF FIGURES 5

LIST OF TABLES 5

LIST OF ABBREVIATIONS 6

1. INTRODUCTION 7

1.1 Background Literature 8

1.1.1 History of Voice Assistants 9

1.1.2 Recent Voice assistants 12

1.1.3 Sinhala Voice Assistants 13

1.1.4 The Voice Assistant System 14

1.1.5 Available Machine Learning Tools for Voice Assistants 16

1.1.6 Advancement of Artificial Intelligence 18

1.1.7 The Sinhala Language 22

1.1.8 Autism Spectrum Disorder 23

1.2 Research Gap 25

1.3 Research Problem 27

1.4 Research Objectives 30

1.4.1 Main Objective 30

1.4.2 Specific Objectives 30

2. RESEARCH METHODOLOGY 31

2.1 System Diagram 38

2.2 System Implementation 39

2.2.1 STT Model 40

2.2.2 NLU model 42

2.2.3 Dialogue Management Model 45

2.2.4 TTS model 46

2.3 Commercialization 47

3. RESULT AND DISCUSSION 48

3.1 Results 48

3.2 Research Findings 52

3.3 Discussions 54
4. STUDENTS CONTRIBUTIONS 56
5. CONCLUSIONS 58
REFERENCES 60
APPENDICES 63

TABLE OF FIGURES

Figure 3.1: Dialogue Management system for ASD……48

Figure 3.2: Speech recognition performance 50

Figure 3.3: Age 4-6 yrs typical speech detection and non-speech detection 50

Figure 3.4: Gender wise ASD in toddlers from 4-6 years 52

Figure 3.5: Age wise ASD detection in toddlers 53

Figure 3.6: ASD disorders sub section in children gender wise 53

LIST OF TABLES

Table 1: Workload Distribution 56

LIST OF ABBREVIATIONS

Abbreviation	Description
ASD	Autism Spectrum Disorder
IVR	Interactive Voice Recognition
OCR	Optical Character Recognition
PC	Personal Computer
AI	Artificial Intelligence
GPU	Graphical Processing Unit
NLU	Natural Language Understanding
NLP	Natural Language Processing
RASA NLU	An Open Source NLU tool
SNIPS NLU	An Open Source NLU tool
HMM	Hidden Markov Model
TTS	Text to Speech
CDC	Center for Disease Control
IDE	Integrated Development Environment
NLG	Natural Language Generator
ICT	Information Communicational Technology
API	Application Programming Interface
ML	Machine Learning

1. INTRODUCTION

The Sinhalese people, Sri Lanka's biggest ethnic group, speak Sinhala as their first language. There are around 16 million Sinhalese people worldwide. Considering that Sinhala is also spoken as a second language by other ethnic groups. Sinhala is a widely spoken language in Sri Lanka. 19 million individuals are affected.

Sinhala is the sole language spoken by the majority of Sinhalese. According to the Department of Census and Statistics of Sri Lanka, about half of urban youth can read an English newspaper in 2007, but this figure is far below 40% in rural regions. In comparison, the country's total literacy rate is 98.1 percent. As a result, Sinhala language computing is in desperate demand. The foundation for this has been laid with the adoption of Sinhala Unicode. However, the quantity of research done in the field of Sinhala Natural Language Processing (NLP) is insufficient. Unlike languages like English, Spanish, or French, which are spoken by a larger number of people throughout the world, Sinhala is only spoken in Sri Lanka. This has a negative influence for Sinhala NLP research. Although some preliminary research exists in areas such as Sinhala-English translation (Silva and Weerasinghe, 2008), Sinhala-Tamil (Sri Lanka's other official language) translation (Sripirakas et al., 2010), and Sinhala spell checking (Jayalatharachchi et al., 2012), the attention paid to processing of spoken and written Sinhala conversations is very low. The goal of this project is to establish the groundwork for filling this gap in the processing of Sinhala dialogues, both oral and written.

Virtual assistants have become a commonplace aspect of our lives. We ask Siri nearly everything we're curious about, and we use Alexa to do grocery orders. Voice assistants are quite useful in corporate applications, in addition to offering ease in our daily life. We utilize internet chatbots to assist us negotiate complicated technological difficulties, make insurance claims, and book hotel reservations, for example. To assist scale client relationship management, we also utilize totally automated calls. To grow to hundreds of millions of end users, each of these apps necessitates the deployment of a production-grade, powerful text-based or voice-based Voice assistant. Voice interfaces have become the primary facilitators for supporting high-quality human/machine interfaces due to their naturalness. Many developers, however, find voice-based virtual assistants to be a substantial technological barrier, especially when deployed at scale.

Autism spectrum disorder (ASD) is a communication and behavior condition that affects children. Although autism can be diagnosed at any age, it is classified as a "developmental condition" since symptoms usually show between the ages of two and two. Autism is referred to be a "spectrum" condition since the kind and degree of symptoms that people experience vary greatly. ASD affects people of all ethnic, racial, and socioeconomic backgrounds. Despite the fact that ASD is a lifelong disease, therapies and services can help a person's symptoms and function.

1.1 Background Literature

In recent years, there has been a notable increase in research aimed at identifying biological and behavioral markers to help in the early diagnosis of Autism Spectrum Disorders (ASD). ASD is a group of disorders characterized by social, language, and communication impairments, as well as stereotyped behaviors. Early diagnosis is critical for better treatment outcomes and less caregiver stress. Autism is known to manifest itself in a number of ways in the communication of children and adults. Common linguistic abnormalities include echolalia, out-of-context words, pronouns, and role reversal.

[1] However, with autism, there are several distinct kinds of verbal abilities. As a result, linguistic-based markers may be ineffective in detecting ASD. [2] Suprasegmental acoustic aspects related to articulation, loudness, tone, and cadence have shown promising results for children's speech, whereas aberrant prosody, which has also been described as a fundamental hallmark of ASD, appears to be ideally equipped for automatic identification rather than aberrant prosody. These auditory components have also been successfully employed in speech-based engagement systems to help children with ASD improve their social skills.

To diagnose autism automatically, machine learning techniques based on auditory and prosodic feature sets have been studied. [3] While studies show that high levels of reliability can be achieved for tasks like differentiating positive phase children from children with ASC, the performance of such systems has been assessed on relatively small datasets, which might contribute to confounding. The limited number of ASD-related datasets currently available is a major impediment to building strong models that are reliable enough for therapeutic application. Nonetheless, because a suitable tool is not widely available, the vision-impaired population of Sri Lanka has a difficult time connecting with technology.

For companies, voice assistants have become the most preferred approach to act as reps. And, from the dawn of AI, one of the most difficult tasks has been creating the ideal voice assistant.

1.1.1 History of Voice Assistants

I. Kenneth Colby, a psychiatrist, invented PARRY, an early voice assistant output, in 1972. It was made with the intention of mimicking a paranoid schizophrenia patient. The rudimentary model that has been created for this voice assistant is based on the behavior of a person with paranoid schizophrenia, such as ideas, conceptualizations, and beliefs. Actual patients and computers running PARRY were evaluated by psychiatrists, and transcripts of the talks were transmitted to another set of psychiatrists to identify which patients were genuine and which were not. They were only able to identify 48 percent of the persons by guessing.

II. ELIZA was created in the early 1990s as a nondirective psychotherapist simulation voice assistant program. It uses clever handwritten templates to produce answers that mirror the user's input utterances. To analyze input phrases, hand-written decomposition rules are employed, which are activated by key words in the input text. ELIZA's ability to comprehend natural language was severely limited.

III. ALICE (Artificial Linguistic Internet Computer Entity) is a software robot or computer with which users may have natural-language conversations. ALICE uses a pattern-matching system that uses depth-first search techniques to identify human input. For the past two years, it has also passed the Turing test. The Artificial Intelligence Markup Language (AIML) was used to build ALICE (AIML). The most recent version of this language is built on the Pandora platform, with the well-known voice assistant "Mitsuku" as the main character.AIML - Example

A simple example on using AIML is shown below.

```
<category>

<pattern>What is your name? </pattern>

<template>My name is Alice</template>

</category>
```

If the user's input sentence matches the sentence between the pattern> brackets, the outputted reply will be the sentence between the template> brackets. In the following example, the "*" (star) sign is used to substitute words. Whatever word comes after the word as in this example will appear in the response at the location specified by the star/> token:

```
<category>

<pattern>I like *</pattern>

 <template>I too like <star/></template>

</category>
```

IV. Eliza is supposed to have spawned Elizabeth. Elizabeth stores her data as a script in a text file, with each line starting with the script command notation. Elizabeth may produce a sentence grammar structure analysis by representing grammar rules with a set of input transformation rules.

In terms of the aforementioned early approaches, they are the primary levels of the voice assistant's history. "Keyword-based" or "template-based" matching was used by the bulk of early voice assistants. In certain voice assistant systems, 'rule-based' or fundamental statistical approaches are also utilized [3].

In some instances, handwritten rules may or may not be useful. Kumar Shridhar, BotSupply's Co-Chief AI Scientist, claims that "A bot answers to inquiries based on pre-programmed rules in a rule-based manner. The given guidelines might be straightforward or complex. When using a rule-based approach, bot construction is relatively straightforward, but the bot is useless when responding inquiries whose pattern does not match the rules it was taught ".

Furthermore, the template matching approach limits the bots' ability to respond only to user inputs described in pre-determined templates. Manipulation of templates for nearly all user inputs is challenging due to the sophistication of today's languages. As a result, there's a strong chance that a bot won't be able to react legitimately or helpfully to most user inputs. Because Sinhala is an agglutinative language with a high rate of affixes or morphemes per word, the earlier approaches are ineffective.

"Keyword-based" or "template-based" matching was used by the bulk of early voice assistants. In certain voice assistant systems, 'rule-based' or fundamental statistical approaches are also utilized [3]. And, for developing conversational bots, neural networks have become the most preferred model. For the English language, many advanced voice assistants have been developed.

As a result of these limitations in presence, it is apparent that previous approaches must be replaced. They have been replaced by neural network-based techniques. When neural networks are utilized as the backbone of conversational modeling, traditional machine learning methodologies are simply used as a supplement.

Traditional rule-based and neural network-based approaches are distinguished in the latter situation by the presence of a learning algorithm. To transform input words into answers, deep learning models employ matrix multiplications and non-linear functions with millions of parameters, rather than hand-written rules.

V. In the history of voice assistants, the transition to messaging systems outside of standalone programs was the next logical step. The most well-known example is SmarterKid, a snarky voice assistant that debuted in 2001. SmarterKid was ahead of its time, bringing us closer to the voice assistant experience we have today. It was a first exposure to voice assistants for many customers, as it brought NLP to SMS networks and AOL Instant Messenger.

In addition to chatting, SmarterKid performed a number of useful tasks, including giving news, weather, stock information, sports scores, and much more. This demonstrated and promoted the capacity of voice assistants to function as smart digital assistants via popular messaging networks. SmarterKid was only one of numerous voice assistants produced by Active Buddy, which was subsequently bought by Microsoft. Active Buddy also offered a range of advertising bots (including one styled after Austin Powers), reframing the voice assistant as a new marketing tool rather than just a conversation companion.

1.1.2 Recent Voice assistants

Because of recent advances in machine learning, the accuracy and usefulness of natural language processing has greatly improved, making voice assistants a viable option for many organizations. This breakthrough in NLP has inspired a rush of new research, indicating that voice assistant efficacy will continue to improve in the coming years. To create a simple voice assistant, a FAQ (commonly asked questions) may be put into voice assistant software. Connecting the voice assistant to the company's business software may expand its capabilities, allowing it to answer more personal questions like "What is my balance?" or "What is the status of my order?"

The bulk of commercial voice assistants rely on technology built by IT behemoths for natural language processing. Among the services offered are Amazon Lex, Microsoft Cognitive Services, Google Cloud Natural Language API, Facebook DeepText, and IBM Watson. Voice assistants are utilized on platforms such as Facebook Messenger, Skype, Slack, Twitter, Kik, WhatsApp, and Viber.

In today's voice assistants, deep learning technologies are employed. To immediately transform input words into replies, deep learning models employ matrix multiplications and non-linear functions with millions of parameters. There are two types of neural network-based conversational models: retrieval-based and generative models. The former simply computes the cosine similarity between the word embedding of the input utterances and the candidate responses from the dataset based on a scoring function, which may be implemented as a neural network [11], or simply computes the most likely answer to the current input utterance based on a scoring function, which may be implemented as a neural network [11].

There have also been approaches that mix the two types of dialog systems by evaluating which of the produced and retrieved responses is more likely to be superior. Three of the most popular voice assistants now are Apple's Siri [4], Google's Assistant, and Microsoft's Cortana [7].

1.1.3 Sinhala Voice Assistants

I. Hettige, B. and Karunananda, A. Voice assistant

The first Sinhala voice assistant was released in 2006, and it was designed to answer basic questions and was not connected to a certain domain. It's a voice assistant prototype that contains a Morphological analyzer, Sinhala parser, Sinhala composer, and lexical dictionaries, as well as a Sinhala language parsing system. This system was built using Java and SWI-Prolog and can run on Linux and Windows. This system is designed as a client-server system, with the server housing all of the resources and engine modules and clients accessing the data over the network. The client-server architecture is used to allow several users to use the voice assistant system at the same time to search for information.

A knowledge identification engine is used to discover the appropriate pattern in order to get the correct response. The system can distinguish the subject, object, and verb in a sentence. The patterns are then saved using the spattern/3 Prolog predicate. The standard Prolog matching and unification approach is used to find acceptable replies

[2]. As a consequence, it is possible to conclude that this research was carried out in order to construct the whole framework that is necessary.

II. GIC Voice Assistant

This is a button-based voice assistant that is available in English and Sinhala and is incorporated into the Government Information Center's website. In this voice assistant, the user can choose from a pre-defined list of button-like options. Because they don't have to write a single query, customers benefit from a considerably faster experience. Users can instead choose a button from a menu of alternatives. The disadvantage of this sort of voice assistant is that users are limited to a small number of options, resulting in a narrow range of topics addressed. The bot user's information objectives are limited, and he or she cannot freely ask questions.

A bot must fulfill two major tasks in order to react to a user's message. The two issues discussed in the following sections of this Literature Review are intent classification and entity extraction. Wit.ai and LUIS are two built-in apis that can do the two tasks listed above. These technologies make it simpler to build voice assistants by fulfilling two key tasks, which are especially important for English-speaking consumers. However, because https calls are slow, utilizing APIs may cause the application to slow down, and users are always limited by the design decisions made for API endpoints. There's also a chance that the libraries will be hacked.

1.1.4 The Voice Assistant System

A voice assistant system should have the following three features:

- Natural Language Understanding

  This section should include a mechanism for the bot to understand what the user is asking for. First and foremost, the bot must be able to tell the difference between when a user asks information from it and when the user provides it to it. The objective of the user inquiry must be captured after it has been identified. In addition, in order to deliver a suitable answer to the user, specific parameters/entities in the user query may need to be obtained.

- Define and design the voice assistant's knowledge base.

  The bot must reply correctly to the user once it has recognized the user's voice. At this stage, the knowledge base comes into play. To acquire the right results for the user, a well-organized knowledge base should be accessible.

- Develop appropriate pattern matching algorithms.

  For a certain request or response to be provided, there may be a specific format that fits into a collection of user utterances. It is most useful in entity extraction from that user's utterance if a specific format can be discovered. For patterns like these, regular expressions are a well-known approach. In NLU frameworks like RASA, regular expressions may also be created.

I. Intent Classification

In the realm of AI, machine learning, and chatbots, intent classification is the process of categorizing a customer's intent by studying their language. A client who writes "How can I discover my order status" into a chat window, for example, is most likely seeking for order status. The computer recognizes the customer's purpose and leads them to an agent or bot who can assist them with their question.

II. Entity Extraction

Entity extraction, also known as named entity extraction (NER), allows machines to recognize and extract things such as product names, events, and locations automatically. Search engines use it to interpret searches, chatbots use it to communicate with humans, and teams use it to automate time-consuming activities like data input.

1.1.5 Available Machine Learning Tools for Voice Assistants

The core capability of a voice assistant should be natural language understanding (NLU). The following tools/platforms are included as part of Intent Classification and Entity Extraction. Only a few open-source Natural Language Understanding Tools are free, such as RASA NLU and SNIPS NLU, whereas the majority of the others require a membership for more sophisticated capabilities.

I. RASA STACK

RASA Stack's machine learning frameworks may be used by developers to create contextual AI agents and voice assistants that go beyond simple inquiries. Thousands of people in the community are advocating for open-source natural language processing and dialogue management [18].

The two main components are RASA Core and RASA NLU. RASA Core is a voice assistant framework with machine learning-based conversation management, while RASA NLU is a natural language understanding library with intent classification and entity extraction.

These two components are self-contained. Both RASA Core and RASA NLU may be used with a variety of NLU frameworks and dialog management frameworks and tools.

When a user message is received, RASA NLU is in charge of interpreting it based on past training data. To do so, two approaches are used: intent classification and entity extraction. The act of determining meaning based on stated intentions is referred to as "intent classification." The intent returned with the greatest confidence rate is used to assess the intention of a user's utterance. Entity Extraction is a method of locating structured data (the entities and their values).

Following that, RASA Core is activated. The core determines what occurs next in the conversation. Its machine learning-based dialogue management predicts the next best action based on NLU input, conversation history, and training data. This is the research platform; therefore, it will be explored in more depth later.

II. SNIPS NLU

Snips NLU is an open-source Python package for natural language understanding that allows you to extract structured data from natural language phrases [6]. After being correctly trained, the Snips NLU engine can extract structured data. The same method as RASA NLU may be used with this. This includes determining the objective and extracting the entities. It may be used with both pre-built and bespoke entities.

However, the reason SNIPS NLU cannot be utilized in our study is because it requires language resources for the language we is using, and there are presently no language resources available for Sinhala.

III. NLULIte

NLUlite is a developer-friendly database that scans texts and responds to queries using a natural language parser and a graph database [15].

IV. SyntaxNet

SyntaxNet is a TensorFlow toolkit for deep learning-driven natural language understanding (NLU) [16].

V. DialogFlow

Dialogflow is a Google-owned company that creates conversation-based human–computer interaction technologies using natural language. The company is best known for creating the Assistant, a virtual assistant for Android, iOS, and Windows Phone that does tasks and responds to questions using natural language. This NLP framework includes a powerful natural language understanding (NLU) engine for processing and comprehending natural language input, which enables conversational interfaces to be developed on top of products and services [9].

VI. LUIS

The Language Understanding Intelligent Service (LUIS) from Microsoft Azure makes it simple to include language understanding into applications [10].

1.1.6 Advancement of Artificial Intelligence

Artificial intelligence voice assistants with traditional text-based interfaces have become a new phenomenon on the market in recent years. Despite this, the absence of human feeling is a significant issue associated with voice assistants. Due to their lack of human emotion, voice assistants look odd and inconvenient. Voice-activated voice assistants and devices entered the market as a reaction to this problem, with virtual assistants like Amazon Alexa, Google Assistant, and Apple Siri becoming highly popular. Most voice-enabled voice assistant frameworks are made up of the following components.

- Speech Recognition
- Natural Language Processing
- Conversational Artificial Intelligence (Dialogue Management)
- Speech Synthesis

Researchers have conducted a number of research in the aforementioned areas in order to obtain the best possible results in each application. The following are some of the most significant breakthroughs in the disciplines mentioned above.

I. Speech Recognition

The main objective is to educate a computer to understand spoken language. Acting appropriately and translating the receiving speech into another medium, in this case writing, is what understanding entails. Speech recognition is referred to as "voice-to-text" (STT). Despite the fact that a significant number of researchers claim to be investigating this topic at the time, it has been around since 1920, when machine recognition was first presented. Since then, engineers and scientists have experimented with a number of techniques and patterns, which have developed over time. Here are a few examples:

- Hardware-based voice recognition

- Acoustic phonetics-based speech recognition

- Pattern-based recognition of voice
- Speech recognition with continuous word recognition
- To recognize speech, a combination of statistical and connectionist methods (HMM/ANN) is employed.
- The Variational Bayesian (VB) estimate is used.

Some of the ongoing research initiatives are listed below. Speech analysis is conducted in the telecommunication terminal, while recognition is performed at a central point in the telecom network, according to [9] The Aurora framework, which is defining standards for Distributed Speech Recognition (DSR). Several front-end feature extraction ideas are now being compared using the framework [3]. Furthermore, an empirical comparison of the CTC, RNN-Transducer, and attention-based Seq2Seq models for end-to-end speech recognition has been suggested [10].

II. Natural Language Processing

Natural language processing research has gotten a lot of attention in recent years. Natural language processing (NLP) is a tremendously active area of study and development since it is a computerized technique to interpreting text. [6] Paraphrase that has been formalized Once upon a time; research suggested a platform that combined deep learning approaches with GPU-based mind training. The study focuses on basic semantic issues, such as efforts to generalize semantic role labeling to all words, models for generic coreference resolution, semantic parsers that produce relatively competitive meaning representations, and semi-supervised learning of better word representations. In contrast to the foregoing, research [7] proposes a detached natural language processing (Rasa NLU) from the dialog management unit. Its API is built on sckit-learn and Keras, with a consistent API taking precedence over strict inheritance. For text classification, Rasa NLU employs a fastText method that combines pre-trained word embeddings with trained intent classifiers. When compared to competing systems, Rasa NLU comes out on top in benchmarks [8].

III. Conversational Artificial Intelligence (Dialogue Management)

Traditionally, conversation management systems have been designed using a unique pipeline that combines distinct modules for language interpretation, state monitoring, action selection, and language creation. With the method described above, each module must be trained separately with labeled data. However, the engineering process's complexity and closely linked module dependencies compelled researchers to seek alternate alternatives. Researchers were able to build methods that infer a latent representation of the state thanks to recent advances in recurrent neural networks, but they lacked a generic way to inject domain information and restrictions. By [1] training a recurrent neural network on text transcripts of discussion, a latent representation of state may be inferred, eliminating the requirement for state labels. Furthermore, rather than enabling the dialogue management system to acquire the domain knowledge for the scenario from the conversation, [1] provides a method in which the developer has the ability to represent the domain knowledge for the scenario using software and action templates. This method promotes concern separation by allowing domain information and restrictions to be embodied in software and control flow to be learnt from these inputs.

It gives developers more control and only requires a small amount of data to train the system. It outperforms the performance of simply learnt models and rule-based systems, according to the findings of the aforementioned study. Integration of reinforcement and supervised learning in the existing approach is also recommended as a way to improve the system.

According to [2,] present neural model-based conversation production appears to be a tried-and-true approach, with the sole drawback being that the dialogue generating process entirely ignores the discussion's future conclusion. The aforesaid issue was addressed using reinforcement learning algorithms, which allowed produced conversations to take into account the response's future result via a specified reward function. Researchers were able to enhance reward-based interactive answers that encourage a more prolonged discussion using deep reinforcement learning, as mentioned in the above study paper.

[3] The addition of a transfer learning technique to an existing goal-oriented conversation system in a closed domain can increase the system's learning rate by 5 to 10 times, with a response generation success rate of more than 20%. This method, as opposed to a deep reinforcement learning strategy, may be utilized in a space with a limited volume of data. This technique also greatly increases the system's success rate, even when large amounts of domain-specific data are available. The majority of the study was split between flat reinforcement learning agents and rule-based agents for dialogue management.

The latter technique, which is based on a hand-coded approach, relies on a deterministic set of rules and lacks a high-level description of a conversation system [4]. Complex tasks [5] are formulated in a mathematical framework of options over Markov Decision Processes (MDPs), and a hierarchical deep reinforcement learning approach to learning a dialogue manager that operates at various temporal scales outperforms a flat reinforcement learning agent and rule-based agents significantly.

This study introduces a dialogue management engine with three components: a top-level sub-task selector that chooses a subtask or option for the given input, a low-level dialogue policy that chooses primitive actions for the above-selected sub-task, and finally, a global state tracker object that spans the entire conversation tree to keep track of the dialogue's future outcome. According to the research's test results, the agent's hierarchical structure increased the discussion flow's coherence.

IV. Speech Synthesis

Each spoken word is formed by combining a set of vowel and consonant speech sound components in a phonetic manner. Voice Synthesis is the technique of turning any arbitrary text in any language into a matching speech sound unit [11], [12]. Text-To-Speech is another name for this (TTS). Many academics select this issue for their studies since it is now a hot topic in the world of information technology. However, this is an issue that has been discussed in this field since the 18th century [13]. However, scholars in this sector are more interested in the machine learning method. Many different forms of study have been conducted in this subject since its inception, using a variety of languages. Slovak [14], Indian languages - Tamil, Hindi, Malayalam, and Telugu [15], Devanagari script Indian languages [16], Indonesian [17], Moroccan Arabic [18], and so forth.

Based on the objective of their research and the language they concentrated on, they employed a variety of methodologies and procedures. The primary goal, however, remains the same: to transform a text format into a sound signal. There are a variety of speech synthesis methods available, such as [12].

- Unit Selection Synthesis: This approach makes use of a vast database of recorded words, which gives the product a more genuine feel.
- Diaphone Synthesis: This approach maintains tiny units of speech and uses a smaller database than unit selection. As a result, the product is less natural than that produced by unit selection synthesis.
- Domain Specific Synthesis: This approach is typically employed in systems that require a limited vocabulary.
- Formant Synthesis: The source-filter model is used in this approach. Cascade and parallel structures are the two sorts of structures. These two kinds can also be combined to improve performance.
- Articulatory Synthesis: This is based on human speech production system modeling and is difficult to execute. [12], [19]
- Hidden Markov Model

Researchers shifted to a context-independent method since the input is unpredictable when employing voice synthesis. Even humans can learn to pronounce new words quickly by studying the pronunciation of existing ones. This is the foundation of this context-agnostic method.

1.1.7 The Sinhala Language

The Sinhala language dates back over two thousand years. It is a north Indian language similar to Hindi, Bengali, and others. Its most closely related language is Divehi, which is spoken in the Maldives islands (Pannasara and Arachchi, 2011). Con-temporary

Pali, Sanskrit, Tamil, Portuguese, Dutch, and English are only a few of the languages that have impacted Sinhala. The Sinhala alphabet is a Brahmic family script that is employed in the Sinhala writing system. It is one of the world's longest alphabets. Sinhala is one of Sri Lanka's official languages and the mother tongue of 74 percent of the country's people. Sinhala has 40 segmental phonemes, 14 vowels, and 26 consonants in spoken form.

In Sinhala, there are four nasalized vowels that appear in two or three words. They are /ã/, /ã:/, /ǣ/ and /ǣ:/. /æi/, /iu/,/eu/, /æu/, /ou/, /au/, /ui/, /ei/, /oi/ and /ai/ are all diphthongs in spoken Sinhala.

1.1.8 Autism Spectrum Disorder

ASD is a complicated developmental disease characterized by chronic difficulties with social communication, limited interests, and repetitive conduct. While autism is a lifelong condition, the degree to which these problems affect one's ability to operate differs from person to person with autism [19].

Before a kid turns one year old, parents/caregivers or physicians can detect early indications of this condition. However, by the time a youngster is two or three years old, symptoms are usually more persistent. In certain circumstances, the functional impairment associated with autism may be modest and not noticeable until the kid begins school, after which their impairments may become more evident when they are with their classmates [19].

Autism spectrum disorder (ASD) is a social interaction, communication, and behavior problem. ASD can be detected at any age, although symptoms usually appear within the first 24 months of life. In low- and middle-income countries (LMICs) like Sri Lanka, however, there is little evidence supporting ASD prevalence estimates. Due to research and financial constraints, ASD diagnosis in LMICs is lower than in developed nations. Sri Lanka has only conducted little study to establish how many of its inhabitants are autistic, and health authorities who claim ASD is not present in the region are likely unaware of how to recognize it. This is a really serious problem, and additional assistance and services for these people in this country are desperately needed. Early detection of ASD in kids allows for aggressive intervention prior to the completion of neural pruning.

I. Symptoms of Autistic Kids

Autistic kids may have difficulty relating to and communicating with others. They may acquire language more slowly, have no language at all, or have substantial difficulties understanding and using spoken language. They may not utilize gestures to compensate for their difficulties with language. Autistic youngsters typically talk to ask for something or to express their dissatisfaction. For social reasons, such as sharing information, they are less inclined to communicate. They also have a hard time recognizing when and how to interact with others in a socially acceptable manner. They might not establish eye contact or allow another person to have a turn in a discussion, for example.

Kids must be able to understand what others say to them (receptive language), express themselves using words and gestures (expressive language), and use their receptive and expressive language abilities in socially appropriate ways in order to communicate effectively.

II. Autism Screening Tools

The advent of technology allows for a variety of methods to autism screening techniques, both official and informal. These might be anything from casual observations to official evaluations. The following are some of the most widely used autism screening tools:

- The M-CHAT (Modified Checklist for Autism in Kids, Revised) is a common 20-question exam for toddlers aged 16 to 30 months. According to new research, the M-CHAT may be less successful in screening females, minorities, urban youngsters, and kids from low-income families.

- The Ages and Stages Questionnaire (ASQ) is a developmental screening instrument that looks at developmental issues at different ages.
- STAT (Screening Instrument for Autism in Toddlers and Young Kids) is a twelve-activity interactive screening tool that evaluates play, communication, and imitation.
- PEDS stands for Parents' Evaluation of Developmental Status and is a developmental parent interview that looks for deficits in motor, language, self-help, and other areas.

1.2 Research Gap

There is a knowledge gap. A subject or sector in which a lack of or inadequate knowledge makes it difficult to make a conclusion on an issue. What's the link between a research gap and research constraints? The present research literature has a huge void in it. It's the gap that your research approach attempts to fill. To identify gaps, a comprehensive literature study and review is conducted. Significant research has been conducted in the following areas in connection to the potential solutions mentioned in this article: speech recognition, interactive AI, and voice synthesis.

Many of the experiments looked at resulted in the product being able to do the tasks listed below, including the one proposed in this paper. The platform now has multilingual capabilities for real-time speech processing thanks to improvements in speech recognition technology. According to research conducted on each system, the majority of systems do not enable speech to text. Currently, a speech synthesis system's user is limited to the platform's voices, which are generally sent to listeners right away.

It's tough to find a helpful platform to help you improve your voice for a more natural tone after a trial run. Despite the fact that the bulk of these platforms span languages spoken all over the world, we were unable to discover a Sinhala-based platform. [30]

This sort of research has been enhanced by the ability to manage the present state of a conversation, which has replaced the previous unstable dialog control with a state dialog. Thanks to improvements in GPU setup and machine learning, we were able to focus on deep learning-based neural network access for conversational control research. In recent research in this subject, this approach has been employed widely. The issue of supporting the Sinhalese language is still significant in this field of research.

Speech Recognition, Conversational Artificial Intelligence, and Speech Synthesis, all of which are linked to the suggested solution presented here, have all seen a considerable amount of research. Many studies have shown that the products that are similar to the approach provided in this study can perform the following tasks.

Systems can now perform real-time voice recognition in a range of languages thanks to improvements in speech recognition technology. According to the study done on each platform, the majority of platforms do not allow the voice recognition method. Users of currently available speech synthesis platforms are limited to the voices provided by the platform, which makes the bulk of them sound like automated presences to the listener. It was difficult to find a framework that enabled the training of distinct voices to provide more lifelike voice output following the process. Despite the fact that most of these platforms accept languages from all over the world, we have yet to identify a single platform that supports Sinhala.

In connection to Dialogue Management, researchers have utilized a number of approaches to achieve the end goal of task completion, with the majority of them focusing on the English language. Some studies concentrate on developing an initial knowledge base that may be used to map talks using entities and actions. The output is generated by mapping the input source content into knowledge base entities and actions in the respective systems.

These types of research now include the ability to govern the current state of the conversation, changing classic stateless dialogue management into stateful dialogue management. Researchers were able to focus on a deep learning-based neural network approach for dialogue management because to the advancements in GPU processing and machine learning. The most recent study in this field has focused on this method. In these research areas, there was also a problem with Sinhala language support.

This platform, on the other hand, will primarily focus on voice detection, conversational AI, and Sinhala speech synthesis. The platform is a self-contained system that can be implemented locally and customized to meet the organization's needs. Furthermore, unlike most other platforms, the machine learning algorithms are trained using current data in order to generate a domain-specific representation.

1.3 Research Problem

Autism spectrum disorder (ASD) is a brain illness that can cause significant social, communication, and physical problems. The "spectrum" refers to the range of disorders, skills, and levels of disability that children with Asd can have. People with ASD experience a range of symptoms, from mild to severe. Children with Asd have certain problems, such as difficulty with social contact, but there are differences in when symptoms start, how severe they are, the range of indicators, and whether or not other disorders are present. Clinical signs and symptoms might change over time.

The intellectual and spiritual development of children with ASD determines their ability to communicate and utilize language. Some children with ASD may be unable to communicate due to delays in speech and language development, while others may have limited speaking ability. Others may have large vocabularies and be able to speak in depth on a variety of subjects. The meaning and rhythm of words and phrases are difficult for many individuals to grasp. They may also have difficulty reading body language and understanding the implications of different voice tones. When these difficulties are considered together, they have an influence on the ability of children with ASD to interact with others, particularly peers their own age.

Every day, technology advances, making people's lives easier. Because they are widely available and frequently utilized, smart gadgets are an excellent platform for a computer-aided tool. As a consequence of consumers' curiosity and demand for voice assistants, this fact may attract a lot of attention. A voice assistant is a built-in software program that allows users to speak in natural language with one another. Voice assistants are used for a variety of purposes, but their fundamental objective is to detect and respond to the words spoken by users.

Voice assistants are used in e-commerce, insurance, health care, retail hospitality, and logistics, to name a few industries. The most frequent business functions that voice assistants can do include user assistance, sales and marketing, order processing, and social networking [13]. Frequently Asked Questions (FAQ) in a certain subject are often programmed into voice assistants.

It's crucial to keep note of voice assistant encounters since they can help you measure your development. Users can also provide good or negative comments on their experience. The feedback is useful in shaping future decisions. If correctly designed, voice assistants have the potential to eliminate repetitive tasks, making workloads lighter, easier, and faster while also boosting user happiness in businesses that employ them.

Furthermore, if users must wait a long time for a bot to react, their voice assistant experience becomes less beneficial, and they may grow irritated. As a result, whatever internal processing is necessary to interpret the user's message and provide an appropriate response, the user's waiting time should be kept to a minimum. If this does not happen, users may get unhappy. Advances in the fields of Machine Learning and Artificial Intelligence have led to the development of advanced voice assistants.

Due to the fact that most voice assistants are only available in English, some users may be unable to fully utilize the existing voice assistants due to a language barrier. The amount of difficulty in communicating reduces since users are not forced to participate in a formal dialogue while conversing with a bot in their native language. Users may be cautious to ask questions if this is not the case, stopping them from openly conversing. Furthermore, no voice assistants for the Sinhala language have been created to target a specific domain.

A generic voice assistant has a high probability of failing to respond to the majority of questions, rendering it worthless. Instead of wasting time searching the internet, users may ask questions and solve problems by pointing to a specific location when utilizing a domain-specific voice assistant. As a consequence, developing "A Sinhala voice assistant for screening children with autism spectrum disorder" largely solves two concerns by being language and domain specific (Sinhala). However, with only the dataset and a few custom components changed to match the domain, this approach should work in any domain.

Individuals tend to use simple terms in most conversational venues. As a result, the issue of linguistic complexity in voice assistant systems, such as in lengthy conversations, is less of a concern. The Sinhala voice assistant can also take use of this chance. Furthermore, if the bot's ability to discern the user's goal is hampered by the user's speech, the bot might ask the user to repeat what they've previously said.

The discussion grows more lifelike by increasing the user's participation until he or she receives an appropriate response. Furthermore, voice assistants do not need to be concerned about a language's sophisticated written grammatical structures because basic verbal grammar suffices in most cases [2].

The two main types of bots are retrieval-based bots and generative-based bots. A retrieval-based bot uses a prepared selection of responses and a heuristic to pick the best suited response for a user input. A generative model, unlike a retrieval-based bot, aims to turn a given user input into an output by generating a suitable answer rather than utilizing a pre-programmed set of responses.

An increasing number of businesses are experimenting with incorporating a voice assistant into their daily operations. Voice assistants may be utilized to automate various operations in sectors such as medical, education, information retrieval, business, e-commerce, and entertainment, according to recent studies. [14].

Every day, technology advances, making people's lives easier. However, there are certain areas that have yet to be investigated. Due to cultural factors, ASD awareness is low in LMICs like Sri Lanka. People with ASD are typically left untreated for long periods of time after being identified due to a lack of resources. Early detection and diagnosis are essential for improving the clinical outcomes of young babies with ASD. Because they are widely available and frequently utilized, smart devices are an excellent platform for a computer-aided tool.

The bulk of screening systems on the market today automate simple screening checklists like M-CHATR/F. Only the ASD allows you to make a decision. An intelligent machine learning model is used in the AI software. This study presents a novel technique for ASD screening that incorporates a culturally relevant symptom checklist and an integrated machine learning algorithm. A variety of supervised learning models were trained using PAAS data obtained clinically. In terms of prediction, the proposed application beat traditional paper-based approaches (PAAS). The new program aims to promote ASD awareness and detection by allowing non-specialist healthcare practitioners to screen for ASD during home visits.

Furthermore, relevant data on the prevalence of ASD in LMICs (which is currently lacking) may be acquired, and resources could be effectively allocated to decrease treatment delays.

1.4 Research Objectives

1.4.1 Main Objective

The proposed system's aptitude to sustain both English language and Sinhala languages. Research in speech recognition has not sophisticated for Sinhala languages.

1.4.2 Specific Objectives

To reach the main objective specific objectives that need to be attained are as follows.

- Handle countless requests at a time

With the proposed tool. With the proposed platform, incandescent agents will be deployed depending on the current load of requests to be handled and the agents can handle multiple conversations concurrently compared to its human counterpart.

- Escalate altogether productivity of the service.

Intelligent agents formerly deployed by the proposed platform will be reactive and efficient. Due to the incandescent agent’s aptitude to interact with multiple users at a time, users of the entity can maintain in productivity, time, and scalability.

- Enhancing the service's altogether productivity

Intelligent agents are reactive and effective formerly deploys via the advisable platform. Users of the entity can maintain productivity, time, and scalability cheers so the incandescent agent's capacity to convey with many users at once.

- Lowered cost

The entity is simple to configure to fulfill a variety of demands all over time. When compared to already system, incandescent agents will have low to no known maintenance sum formerly installed. Due to the modular foundation no which it was developed, the scheme is straightforward to adapt to several languages.

2. RESEARCH METHODOLOGY

As per the objectives ASD.AI is a complete ASD screening tool. It is a customized voice assistant tool built upon open-source Rasa framework powered by Mozilla. This system includes four basic components. They are the components of speech recognition component, Natural Language Understanding (NLU) component, dialogue management component and speech synthesis component. An image of the components is shown below. The customized system can depend upon existing analysis found in key areas of the platform like speech recognition, colloquial computing, and text-to-speech with the distinctive ability to support the Sinhala language and a standalone, decoupled answer. The answer can set heavily on machine learning algorithms to achieve human-level intelligence in decision-making, generating dialogue, and taking action that may maximize the probabilities of with success achieving the top goal of evaluating patients with ASD [1].

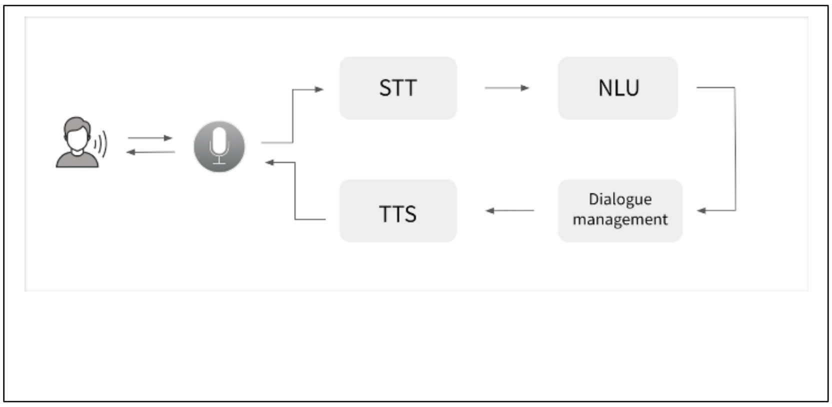


Figure 2.1: Conversational AI Process

The NLP process capability of the projected platform can extend the ASCII text file implementation of the Rasa framework by adding support for Sinhala language process. Once receiving the speech knowledge as text via speech recognition, the tongue process engine extracts the intent, entities and alternative structured data from the Sinhalese language text. The processed knowledge from the tongue process unit is shipped to the dialogue management unit. The dialogue management unit liable for generating answers to queries from users with the last word goal of finishing the task are going to be developed taking under consideration the assorted implementations projected by recent analysis for prime exactitude and speed. The design of the dialogue management engine can closely tally the design projected by the Rasa dialogue management unit. The most focus of victimization the higher than design is that the ability to support machine repair wherever users will answer whether or not the supposed motor action is correct or incorrect reckoning on the state of affairs. These options permit the platform to supply actions to human agents at the initial pace of the platform to assess itself. Additionally, the platform will show a graph of coaching dialogues which will then be become a knowledge domain for the business. The projected dialogue management unit are going to be custom-made by applying deep reinforcement learning with the flexibility to come up with sentences that optimize future rewards, with success capturing the final characteristics of fine speech. This alteration permits for a lot of numerous and interactive responses that promote a lot of property speech [2].

Initially, children's speeches are entered into the system. The system noise suppressor removes noise related to this voice file. Then, the speech recognition engine for the planned resolution was developed supported the deep speech implementation planned by the Baidu researchers, following the newest version of the implementation, wherever version three extracts entities and actions related to knowledge from this voice in text format. The delivered implementation is going to be changed consequently to support Sinhala speech recognition. The key consideration choosing the recent implementation depends on the design instructed by the previous implementation. Instead of ancient text-to-speech engines with intricately designed process pipelines and poor performance in droning environments, the approach taken could be a well-optimized RNN coaching system that uses multiple gpus to coach on an oversized quantity of various knowledge. This offers North American countries the flexibility to coach Sinhala language-specific knowledge while not having to implement custom channels and sound dictionaries for the language. To boot, the implementation emphasizes the flexibility to handle difficult droning environments within the planned system. The chosen resolution even supports coaching through a labelled dataset of transcripts. The machine learning model for speech recognition is going to be trained to support the prevailing body of Sinhalese conversations for the Sinhala language. Additionally, for land language, Mozilla's common voice knowledge set is going to be used [3] [4].

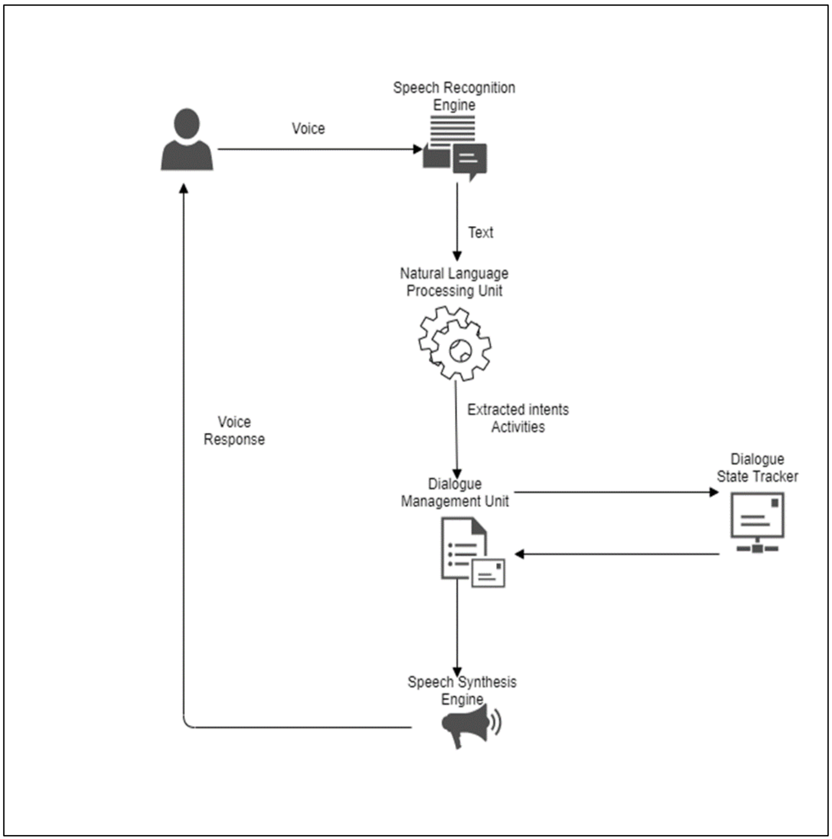


Figure 2.2: System Architecture

Rasa framework is selected to this project as it is an open-source framework, and it makes it easier to implement automated conversational software with machine learning techniques. Rasa provides for the facility to easily implement the customized chatbot and provides the ease of handling and ease of understanding. Moreover, backed by Mozilla, Rasa provides a vast community support. “The Rasa Community is a diverse group of developers, data scientists, designers, and conversational AI enthusiasts [5]”. Further, Rasa framework can be modified for different language support which makes it suitable for this project to customize for Sinhala language. Nevertheless, Rasa is trusted by worldwide brands such as Adobe, Toyota, AIRBUS, BMW, KBC, Allianz, HCA etc. making it a trustworthy platform.

I. Speech Recognition

First, in the speech recognition component, the received speech sequences are translated into text design. The program is implemented based on Baidu's Deep Speech 3 implementation to provide English and Sinhala languages. The recurrent neural networks were trained using existing voice threads using different GPUs. Separate linguistic models will be implemented for each language.

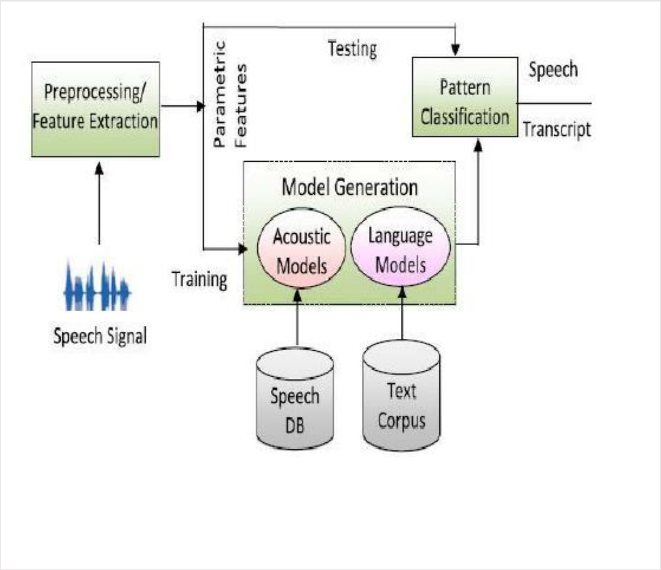


Figure 2.3: Speech Recognition System Architecture

II. Natural Language Processing

Then convert natural language into structured information by extracting intents, entities, and different structured information. Language process modules are enforced supporting each Sinhala and English. A deep learning neural network are enforced to extract necessary keywords like entities, intent and actions from the given text input. The module is enahnced to self learn.

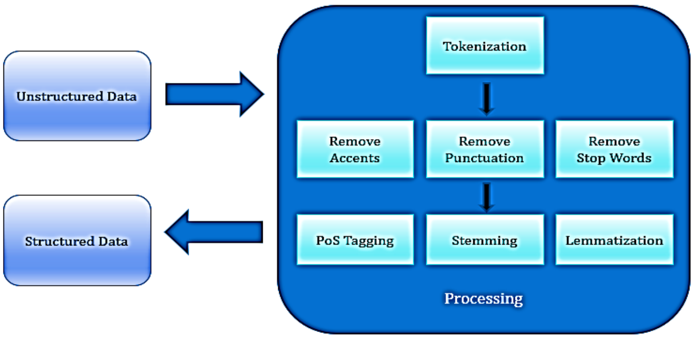


Figure 2.4: NLP System Architecture

III. Dialogue Management

Third, the architecture of the dialog management engine will be closely aligned with the Rasa architecture. The main focus of using the above architecture is the ability to support machine remediation, where users can indicate whether the engine predicted action is correct or incorrect depending on the scenario. These features allow the platform to offer actions to human agents at the platform's initial pace to self-assess. In addition, the platform can display a training dialogue board which can then be converted into a knowledge base to master the system. The proposed dialogue management unit with the flush architecture will be tailored by applying deep learning enhancement with the ability to generate statements that optimize future rewards, successfully capturing the overall properties of a good conversation. This change allows for more diverse and interactive responses that foster a more lasting conversation.

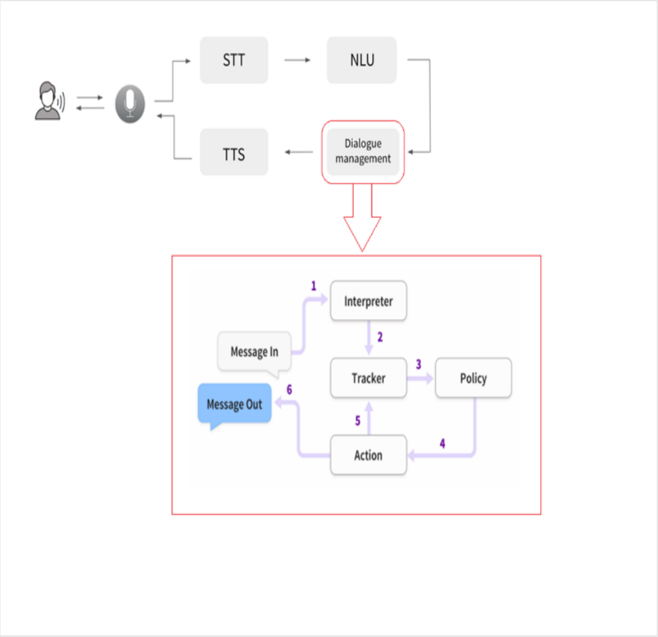


Figure 2.5: Dialogue Management System Architecture

IV. Speech Synthesis

Finally, the component of speech synthesis the system is enforced supported the Deep Voice three and Wave web implementation to support English and Sinhala languages. The continual neural network is trained with a module that contains associate encoder, a decoder and a convertor. The encoder converts the matter options into an enclosed learned illustration. A decoder decodes the learned illustration whereas the convertor creates the audio post-processing to change an additional human voice.

The synthesis part performs the maximization. This can be seen as a reverse operation for speech recognition. First, a given sequence of words is converted into a context-dependent sequence of tags, and then the HMM statement is constructed by concatenating the context-dependent HMMs according to the tag sequence. Second, the speech parameter generation algorithm generates the excitation and spectral parameter sequences from the HMM expression. Although there are several variants of the algorithm for generating speech parameters, most commonly the Case 1 algorithm has been used. Finally, a speech waveform is synthesized from the spectral and excitation parameters generated using excitation generation and a speech synthesis filter. Details about the speech parameter generation algorithm are described below.

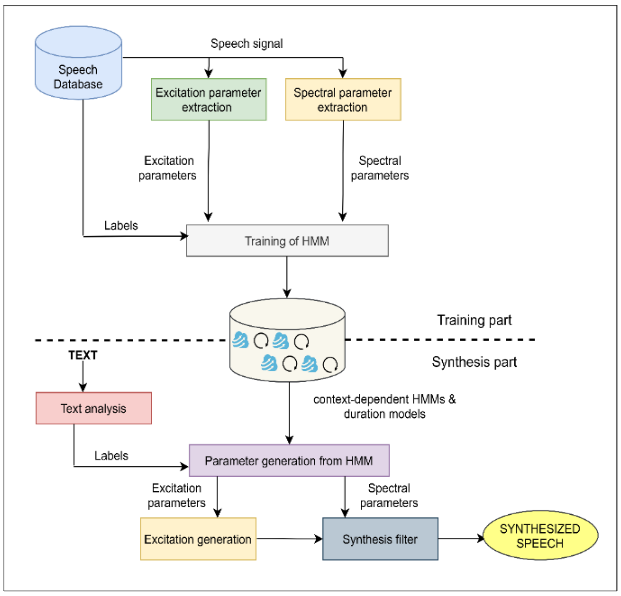


Figure 2.6: Speech Synthesis System Architecture

2.1 System Diagram

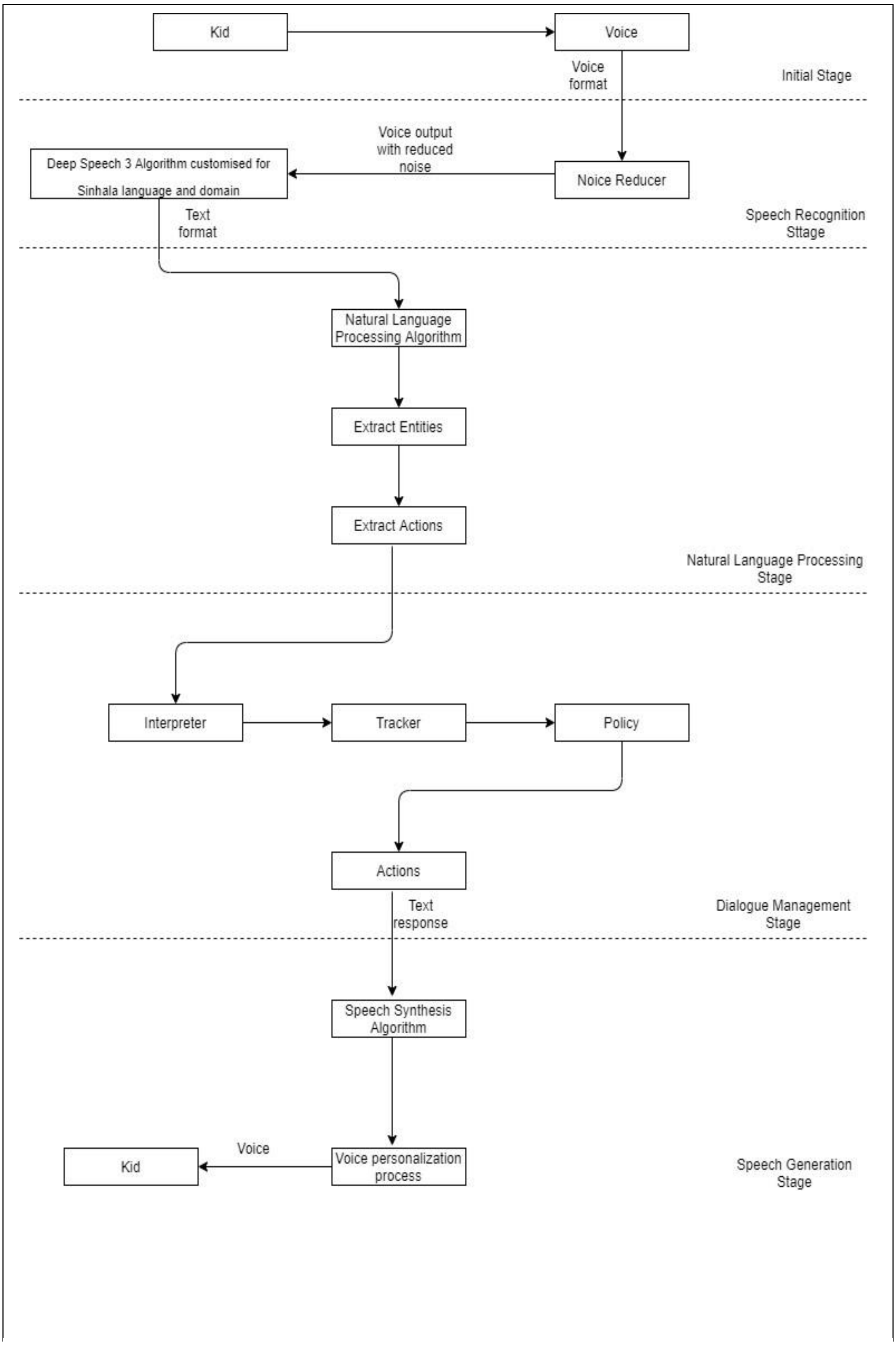
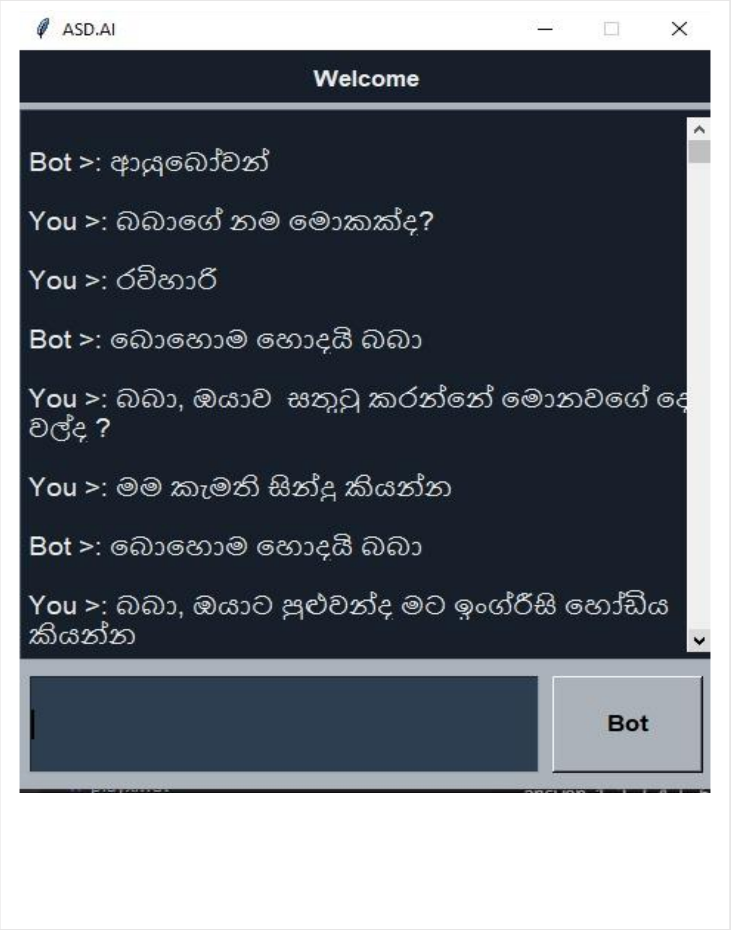

Kid
Voice
Voice format
Initial Stage
Voice output with reduced noise
Deep Speech 3 Algorithm customised for Sinhala language and domain
Noice Reducer
Text format
Speech Recognition Sttage
Natural Language Processing Algorithm
Extract Entities
Extract Actions
Natural Language Processing Stage
Interpreter
Tracker
Policy
Actions
Text response
Dialogue Management Stage
Speech Synthesis Algorithm
Voice
Kid
Voice personalization process
Speech Generation Stage

Figure 2.7: System Overview Diagram

2.2 System Implementation

As shown in the above diagram, there are five basic mechanisms in the framework, namely, voice interface, Speech-To-Text (STT), NLU, dialogue management and Text-To-Speech (TTS). Initial stage denotes the voice interface. Here, this represents a front-end User Interface (UI) which is being used to communicate with the kid. Below figure shows the final output UI in Sinhala language. The chatbot has been implemented to carry out a conversation in Sinhala language. The implementation procedure will be discussed in the next few chapters.



Bot >: ආයුබෝවන්

You >: බබාගේ නම මොකක්ද?

You >: රවිහාරි

Bot >: බොහොම හොදයි බබා

You >: බබා, ඔයාව සතුටු කරන්නේ මොනවගේ දෙ
වල්ද ?

You >: මම කැමති සින්දු කියන්න

Bot >: බොහොම හොදයි බබා

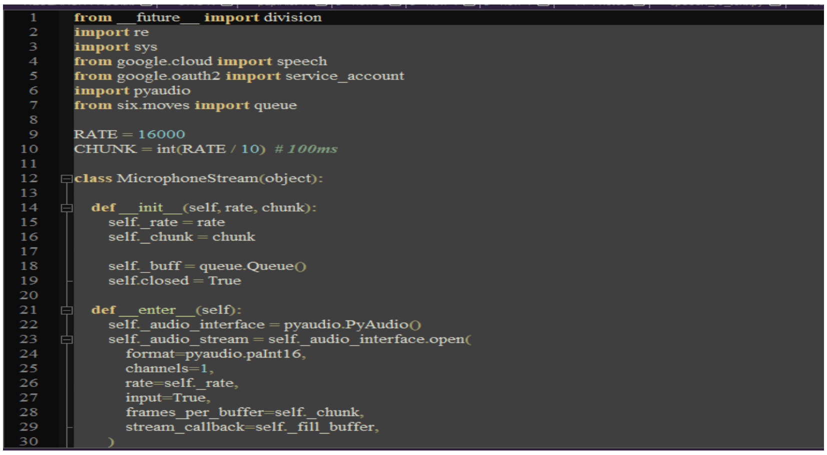
You >: බබා, ඔයාට පුළුවන්ද මට ඉංග්‍රීසි හෝඩිය
කියන්න

Figure 2.8: ASD.AI voice interface

2.2.1 STT Model

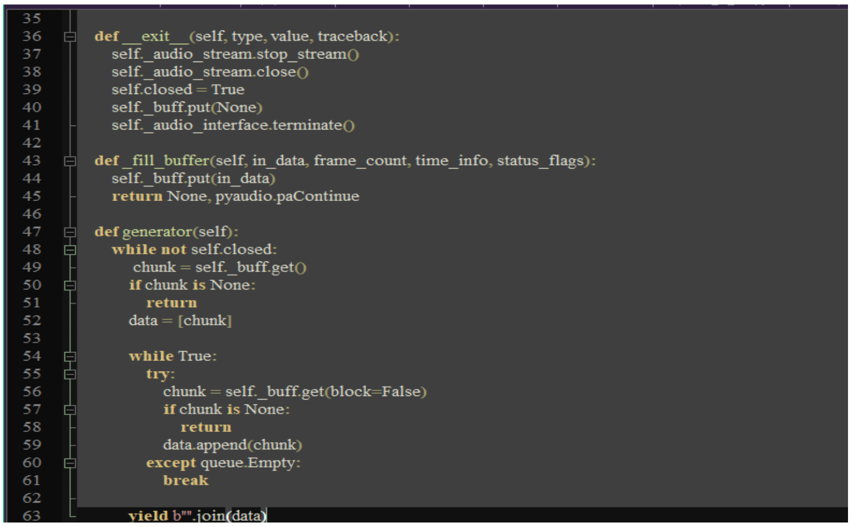
The module has been developed by Mozilla Deep Speech Model. It offers some pre trained libraries with the opportunity to customize. Deep Speech is a deep RNN that is trainable from start to finish, at a character level. That is a deep neural network with recurring layers that takes audio characteristics as input and generates characters directly in the audio transcript. It can be trained using guided learning from scratch, without outside sources of intelligence, such as a grapheme-to-phoneme converter or forced input alignment.

The user input will be taken via the microphone, and it will be passed to speech recognition module.



```python
from __future__ import division
import re
import sys
from google.cloud import speech
from google.oauth2 import service_account
import pyaudio
from six.moves import queue

RATE = 16000
CHUNK = int(RATE / 10)  # 100ms

class MicrophoneStream(object):

  def __init__(self, rate, chunk):
    self._rate = rate
    self._chunk = chunk

    self._buff = queue.Queue()
    self.closed = True

  def __enter__(self):
    self._audio_interface = pyaudio.PyAudio()
    self._audio_stream = self._audio_interface.open(
      format=pyaudio.paInt16,
      channels=1,
      rate=self._rate,
      input=True,
      frames_per_buffer=self._chunk,
      stream_callback=self._fill_buffer,
    )
```

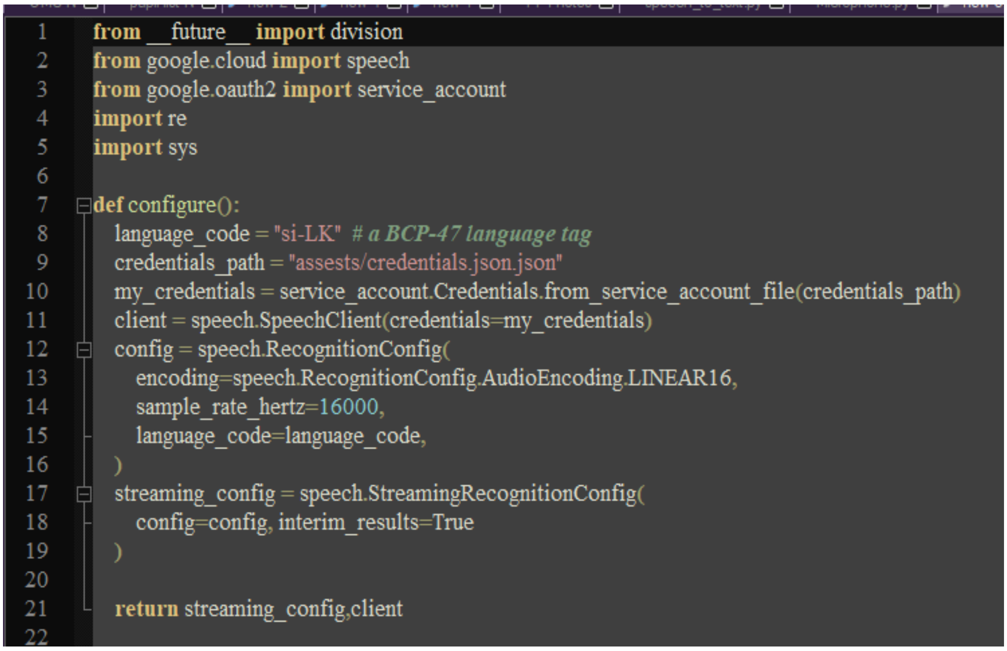
Figure 2.9



```python

  def __exit__(self, type, value, traceback):
    self._audio_stream.stop_stream()
    self._audio_stream.close()
    self.closed = True
    self._buff.put(None)
    self._audio_interface.terminate()

  def _fill_buffer(self, in_data, frame_count, time_info, status_flags):
    self._buff.put(in_data)
    return None, pyaudio.paContinue

  def generator(self):
    while not self.closed:
      chunk = self._buff.get()
      if chunk is None:
        return
      data = [chunk]

      while True:
        try:
          chunk = self._buff.get(block=False)
          if chunk is None:
            return
          data.append(chunk)
        except queue.Empty:
          break

      yield b"".join(data)
```

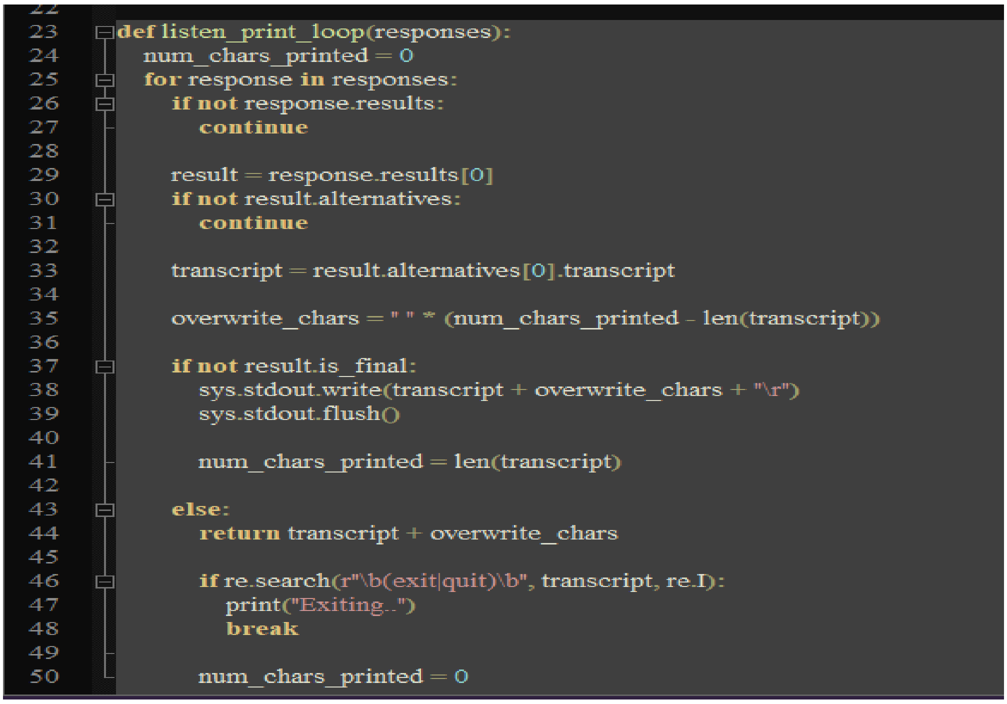
Figure 2.10

Thereafter, the gathered responses will be recognized via ‘streamingrecognition’ module and print an output in text format.



```
from __future__ import division
from google.cloud import speech
from google.oauth2 import service_account
import re
import sys

def configure():
    language_code = "si-LK"  # a BCP-47 language tag
    credentials_path = "assests/credentials.json.json"
    my_credentials = service_account.Credentials.from_service_account_file(credentials_path)
    client = speech.SpeechClient(credentials=my_credentials)
    config = speech.RecognitionConfig(
        encoding=speech.RecognitionConfig.AudioEncoding.LINEAR16,
        sample_rate_hertz=16000,
        language_code=language_code,
    )
    streaming_config = speech.StreamingRecognitionConfig(
        config=config, interim_results=True
    )

    return streaming_config,client

```

Figure 2.11

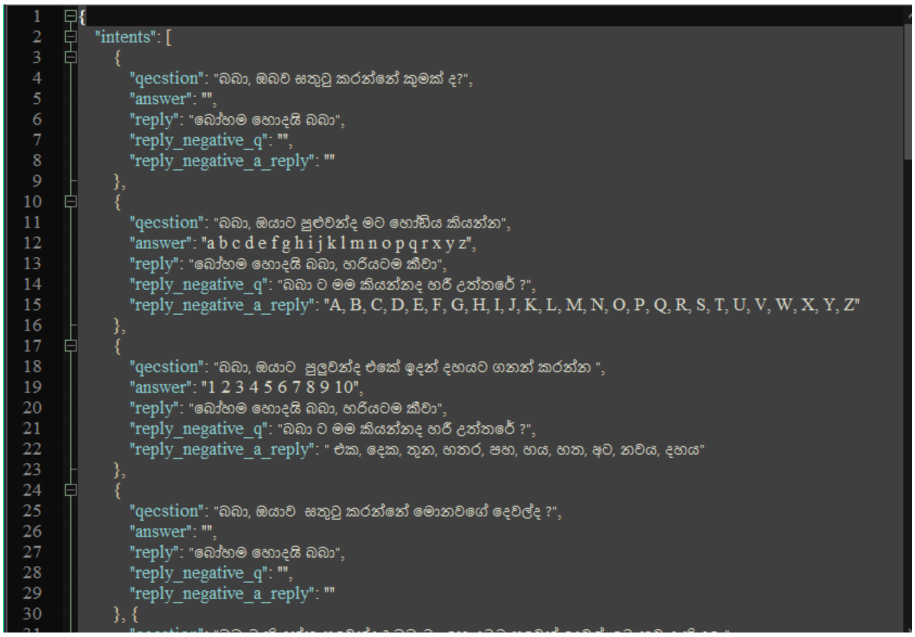


```

def listen_print_loop(responses):
    num_chars_printed = 0
    for response in responses:
        if not response.results:
            continue

        result = response.results[0]
        if not result.alternatives:
            continue

        transcript = result.alternatives[0].transcript

        overwrite_chars = " " * (num_chars_printed - len(transcript))

        if not result.is_final:
            sys.stdout.write(transcript + overwrite_chars + "\r")
            sys.stdout.flush()

            num_chars_printed = len(transcript)

        else:
            return transcript + overwrite_chars

            if re.search(r"\b(exit|quit)\b", transcript, re.I):
                print("Exiting..")
                break

            num_chars_printed = 0
```

Figure 2.12

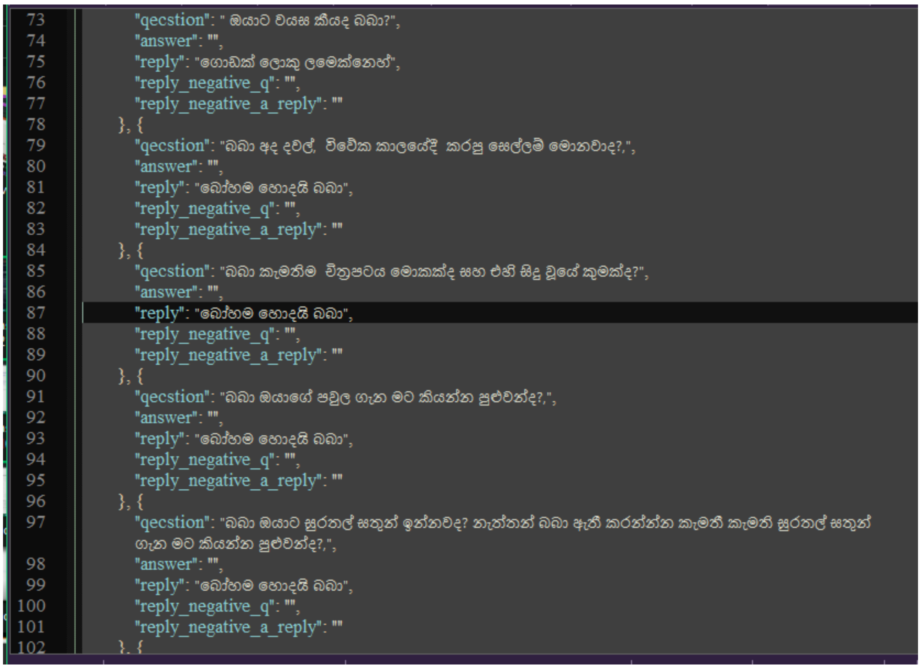
2.2.2 NLU model

In this segment, Rasa NLU has been used for intent classification and for Named Entity Recognition (NER). Firstly, necessary models have been installed. Then a questionnaire as the custom intent file is created using Json format.



```
{
  "intents": [
    {
      "qecstion": "බබා, ඔබව සතුටු කරන්නේ කුමක් ද?",
      "answer": "",
      "reply": "බෝහම හොදයි බබා",
      "reply_negative_q": "",
      "reply_negative_a_reply": ""
    },
    {
      "qecstion": "බබා, ඔයාට පුළුවන්ද මට හෝඩිය කියන්න",
      "answer": "a b c d e f g h i j k l m n o p q r x y z",
      "reply": "බෝහම හොදයි බබා, හරියටම කීවා",
      "reply_negative_q": "බබා ට මම කියන්නද හරි උත්තරේ ?",
      "reply_negative_a_reply": "A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z"
    },
    {
      "qecstion": "බබා, ඔයාට  පුලුවන්ද එකේ ඉදන් දහයට ගනන් කරන්න ",
      "answer": "1 2 3 4 5 6 7 8 9 10",
      "reply": "බෝහම හොදයි බබා, හරියටම කීවා",
      "reply_negative_q": "බබා ට මම කියන්නද හරි උත්තරේ ?",
      "reply_negative_a_reply": " එක, දෙක, තුන, හතර, පහ, හය, හත, අට, නවය, දහය"
    },
    {
      "qecstion": "බබා, ඔයාව  සතුටු කරන්නේ මොනවගේ දෙවල්ද ?",
      "answer": "",
      "reply": "බෝහම හොදයි බබා",
      "reply_negative_q": "",
      "reply_negative_a_reply": ""
    }, {
```

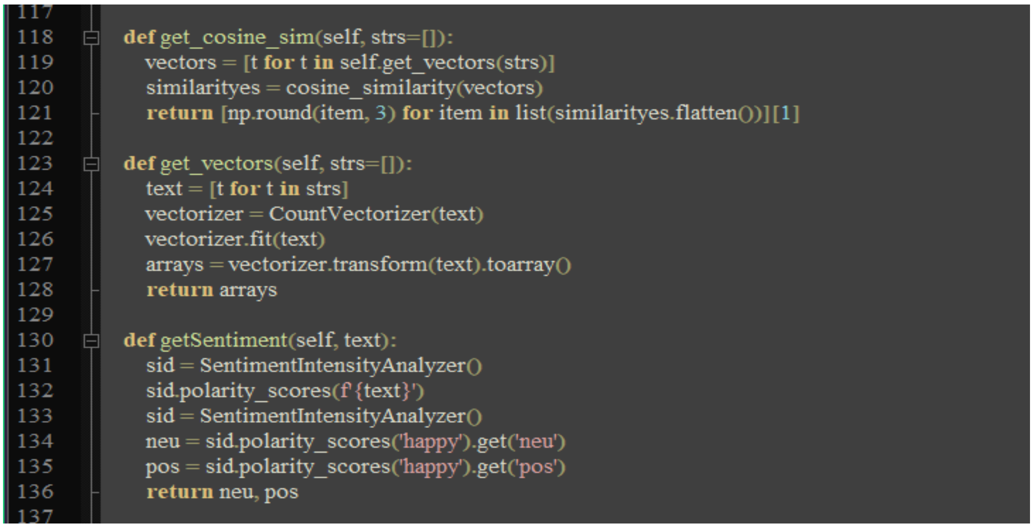
Figure 2.13



```
      "qecstion": " ඔයාට වයස කීයද බබා?",
      "answer": "",
      "reply": "ගොඩක් ලොකු ලමෙක්නෙහ්",
      "reply_negative_q": "",
      "reply_negative_a_reply": ""
    }, {
      "qecstion": "බබා අද දවල්,  විවේක කාලයේදී  කරපු සෙල්ලම් මොනවාද?,",
      "answer": "",
      "reply": "බෝහම හොදයි බබා",
      "reply_negative_q": "",
      "reply_negative_a_reply": ""
    }, {
      "qecstion": "බබා කැමතිම  චිත්‍රපටය මොකක්ද සහ එහි සිදු වූයේ කුමක්ද?",
      "answer": "",
      "reply": "බෝහම හොදයි බබා",
      "reply_negative_q": "",
      "reply_negative_a_reply": ""
    }, {
      "qecstion": "බබා ඔයාගේ පවුල ගැන මට කියන්න පුළුවන්ද?,",
      "answer": "",
      "reply": "බෝහම හොදයි බබා",
      "reply_negative_q": "",
      "reply_negative_a_reply": ""
    }, {
      "qecstion": "බබා ඔයාට සුරතල් සතුන් ඉන්නවද? නැත්තන් බබා ඇතී කරන්න්න කැමතී කැමති සුරතල් සතුන්
      ගැන මට කියන්න පුළුවන්ද?,",
      "answer": "",
      "reply": "බෝහම හොදයි බබා",
      "reply_negative_q": "",
      "reply_negative_a_reply": ""
    }, {
```

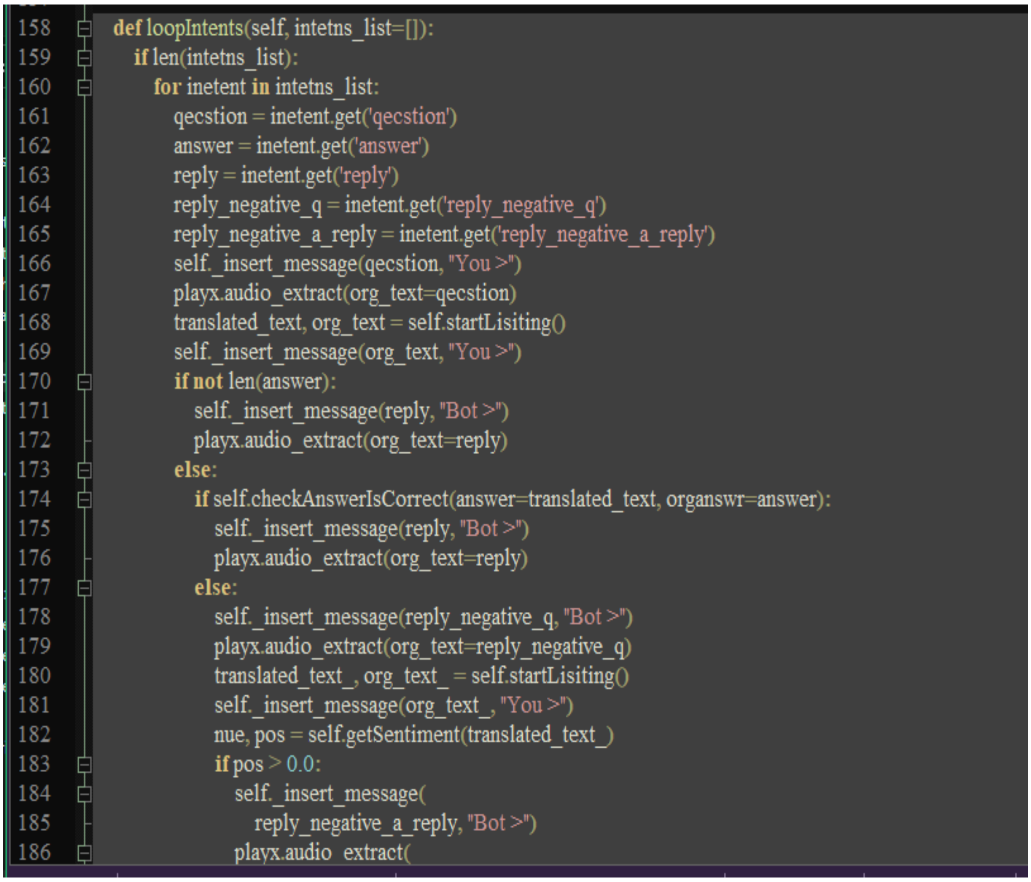
Figure 2.14

These intents are then used in as a loop process in order to carry out the conversation. Moreover, the response data is classified according to their sentiments after the vectorization process.



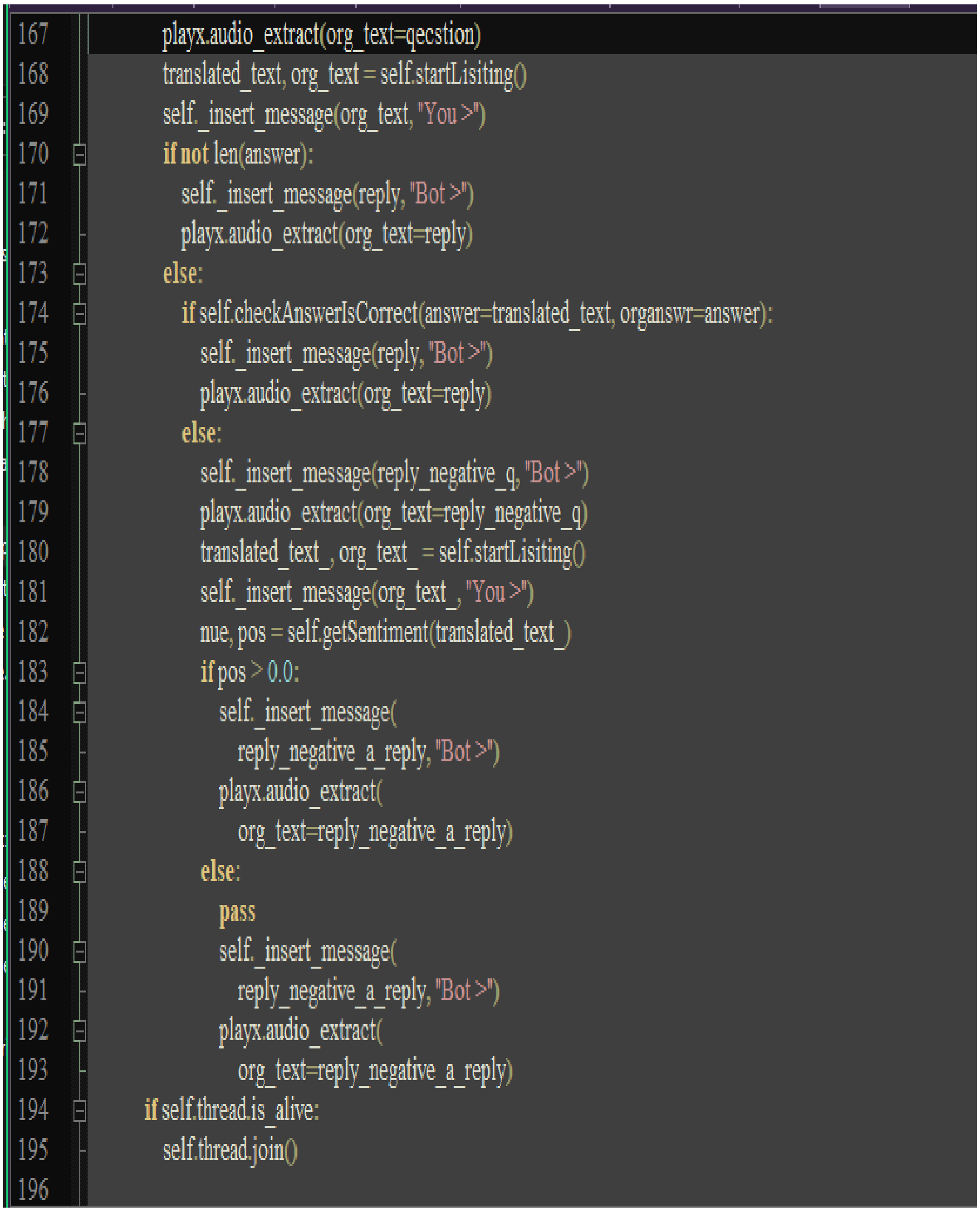
```
def get_cosine_sim(self, strs=[]):
    vectors = [t for t in self.get_vectors(strs)]
    similarityes = cosine_similarity(vectors)
    return [np.round(item, 3) for item in list(similarityes.flatten())][1]

def get_vectors(self, strs=[]):
    text = [t for t in strs]
    vectorizer = CountVectorizer(text)
    vectorizer.fit(text)
    arrays = vectorizer.transform(text).toarray()
    return arrays

def getSentiment(self, text):
    sid = SentimentIntensityAnalyzer()
    sid.polarity_scores(f'{text}')
    sid = SentimentIntensityAnalyzer()
    neu = sid.polarity_scores('happy').get('neu')
    pos = sid.polarity_scores('happy').get('pos')
    return neu, pos
```

Figure 2.15



```
def loopIntents(self, intetns_list=[]):
    if len(intetns_list):
        for inetent in intetns_list:
            qecstion = inetent.get('qecstion')
            answer = inetent.get('answer')
            reply = inetent.get('reply')
            reply_negative_q = inetent.get('reply_negative_q')
            reply_negative_a_reply = inetent.get('reply_negative_a_reply')
            self._insert_message(qecstion, "You >")
            playx.audio_extract(org_text=qecstion)
            translated_text, org_text = self.startLisiting()
            self._insert_message(org_text, "You >")
            if not len(answer):
                self._insert_message(reply, "Bot >")
                playx.audio_extract(org_text=reply)
            else:
                if self.checkAnswerIsCorrect(answer=translated_text, organswr=answer):
                    self._insert_message(reply, "Bot >")
                    playx.audio_extract(org_text=reply)
                else:
                    self._insert_message(reply_negative_q, "Bot >")
                    playx.audio_extract(org_text=reply_negative_q)
                    translated_text_, org_text_ = self.startLisiting()
                    self._insert_message(org_text_, "You >")
                    nue, pos = self.getSentiment(translated_text_)
                    if pos > 0.0:
                        self._insert_message(
                            reply_negative_a_reply, "Bot >")
                        playx.audio_extract(
```

Figure 2.16

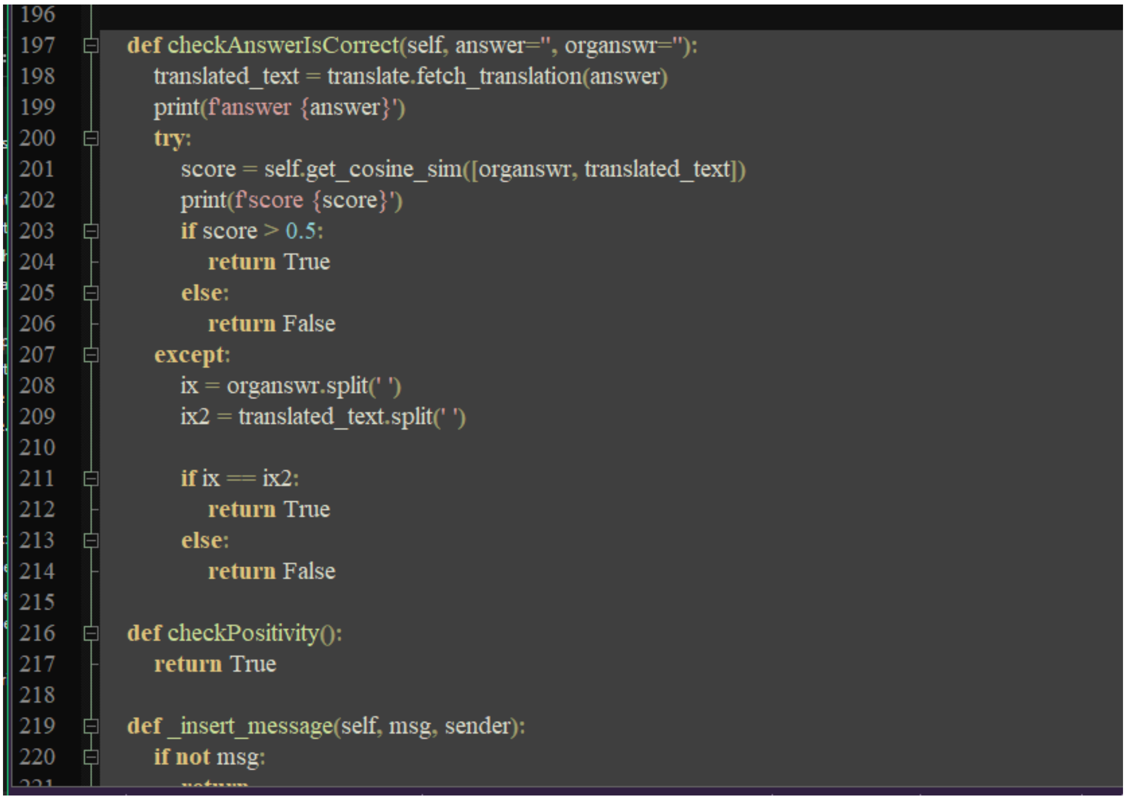


```
        playx.audio_extract(org_text=qecstion)
        translated_text, org_text = self.startLisiting()
        self._insert_message(org_text, "You >")
        if not len(answer):
            self._insert_message(reply, "Bot >")
            playx.audio_extract(org_text=reply)
        else:
            if self.checkAnswerIsCorrect(answer=translated_text, organswr=answer):
                self._insert_message(reply, "Bot >")
                playx.audio_extract(org_text=reply)
            else:
                self._insert_message(reply_negative_q, "Bot >")
                playx.audio_extract(org_text=reply_negative_q)
                translated_text_ , org_text_ = self.startLisiting()
                self._insert_message(org_text_ , "You >")
                nue, pos = self.getSentiment(translated_text_)
                if pos > 0.0:
                    self._insert_message(
                        reply_negative_a_reply, "Bot >")
                    playx.audio_extract(
                        org_text=reply_negative_a_reply)
                else:
                    pass
                    self._insert_message(
                        reply_negative_a_reply, "Bot >")
                    playx.audio_extract(
                        org_text=reply_negative_a_reply)
    if self.thread.is_alive:
        self.thread.join()
```

Figure 2.17

2.2.3 Dialogue Management Model

A training dialog graph can also be displayed by Rasa Core. A story graph is a directed graph that contains nodes that represent actions. The edges of user expressions that occur between the execution of two actions are labeled. If there is no user interaction between two consecutive actions, the border tag is ignored. Every graph has a START node at the beginning and an END node at the end. To simplify the display, a heuristic is used to join similar nodes. During the simplification process, two nodes are merged, creating a single node with all of its properties.

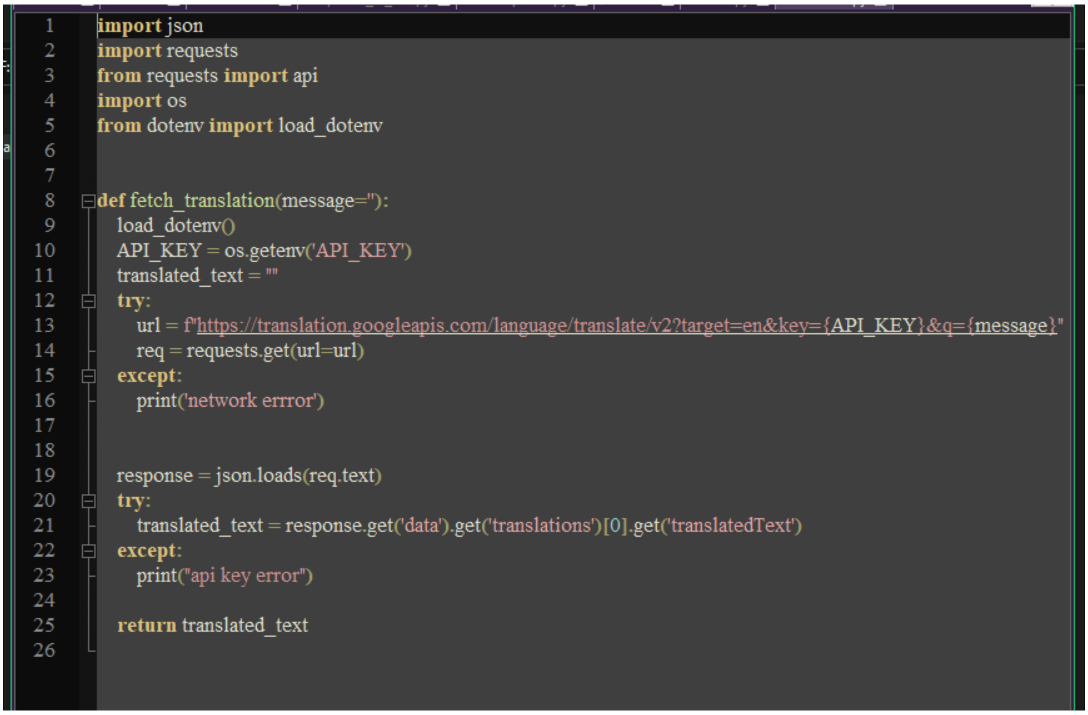


```
def checkAnswerIsCorrect(self, answer='', organswr=''):
    translated_text = translate.fetch_translation(answer)
    print(f'answer {answer}')
    try:
        score = self.get_cosine_sim([organswr, translated_text])
        print(f'score {score}')
        if score > 0.5:
            return True
        else:
            return False
    except:
        ix = organswr.split(' ')
        ix2 = translated_text.split(' ')

        if ix == ix2:
            return True
        else:
            return False

def checkPositivity():
    return True

def _insert_message(self, msg, sender):
    if not msg:
```

Figure 2.18

2.2.4 TTS model

Maximization is done by the synthesis section. The HMM statement is created by concatenating the context-dependent HMMs according to the tag sequence after the given word sequence has been converted to the context-dependent tag sequence. Second, the speech parameter generation algorithm generates sets of excitation and spectral parameters from the HMM expression. Although there are several variants of the speech parameter generation algorithm, the most widely used is the case 1 algorithm. Finally, using excitation generation and a synthesis filter, the speech signal is synthesized from the generated excitation and spectral parameters.



```python
import json
import requests
from requests import api
import os
from dotenv import load_dotenv


def fetch_translation(message=''):
    load_dotenv()
    API_KEY = os.getenv('API_KEY')
    translated_text = ""
    try:
        url = f"https://translation.googleapis.com/language/translate/v2?target=en&key={API_KEY}&q={message}"
        req = requests.get(url=url)
    except:
        print('network errror')


    response = json.loads(req.text)
    try:
        translated_text = response.get('data').get('translations')[0].get('translatedText')
    except:
        print("api key error")

    return translated_text

```

Figure 2.19

2.3 Commercialization

A conversational AI system might cost anything from hundreds to thousands of dollars. A speech pathologist should evaluate an autistic kid to select the appropriate technology for them, program the device with their own language, and teach them how to use it properly. Medicare may cover up to 20 sessions of visiting a therapist about using a speaker system, according to the expert providing the consultation. Some private health insurance policies may pay a portion of the consultation fee. Conversational AI software allows those with visual impairments or reading difficulties to listen to text printed on a smartphone or computer. A visually challenged user can utilize an aural interface to comprehend and execute computer tasks when a conversational AI system is coupled with a screen reader. As a result, this system serves as an assistive device, allowing these persons to use information and communication technology. Governments all around the globe are increasingly looking for innovative methods to support kids with autism spectrum disorder.

Services are expected to account for the bulk of income in the future quarters. The functioning of conversational AI software is reliant on third-party services. They are managed by solution, platform, and service providers and are an integral component of the tool deployment process. Leading companies across a wide range of sectors are employing Conversational AI to cope with the ever-increasing amount of audio/video-based content. This aids organizations in identifying new ways to tap into the huge volumes of data accessible in order to create new goods, services, and processes, giving them a competitive edge.

It used to be simple to contact doctors and medical executives. All that was required of a firm was to send a sales representative to the practice and inform them about the medical devices. Doctors are busy than ever, and many practices no longer let salespeople into their offices. Sales and marketing strategies that worked a decade ago are no longer successful in this industry. Medical device businesses now require new strategies that match how doctors interact with marketing today. Commercial advertising can be used to bring the system in front of a specified audience. This technique may be used to reach a broader audience at various phases of development.

Newspaper advertisements are also effective and may target a wide variety of qualities. Social media platforms may also be utilized to create a digital marketing platform.

3. RESULT AND DISCUSSION

3.1 Results

To reach the aim of the ASD.AI solution, it’s been evaluated among kids with challenges of speaking from different parts of the country who are in the age 5-6 group. One hundred fifty children have been engaged in the voice information gathering, and it’s been acquired from non-autistic youngsters in Sri Lanka are supplied as training data. This dataset includes the characters and words accepted for the literacy contexts, a combination of two lines of text, three-letter phrases, and four-letter words. In the dataset gathering, we’ve been attentive to apply terms that can be comprehended by kids in age 5-6 grouping. When considering about the dialogue management component our system's main intention is to provide live and post-session feedback, as well as to host a dialogue manager that allowed users to engage an open-ended discourse. A Hidden Markov Model (HMM) is used by the real-time feedback system to determine when an icon should become red and then green. The HMM was trained with data from a prior work, and specifics on its algorithm is described in under the methodology section.

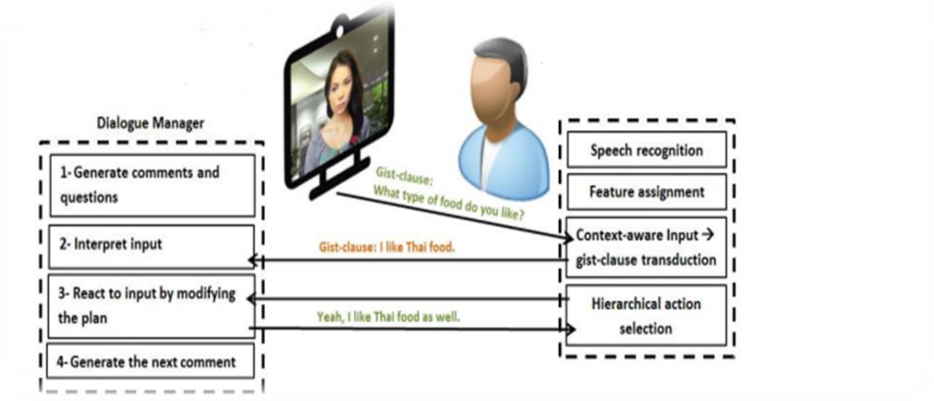

Dialogue Manager
1- Generate comments and questions
2- Interpret input
3- React to input by modifying the plan
4- Generate the next comment
Gist-clause: What type of food do you like?
Gist-clause: I like Thai food.
Yeah, I like Thai food as well.
Speech recognition
Feature assignment
Context-aware Input → gist-clause transduction
Hierarchical action selection

Figure 3.1: Dialogue Management system for ASD

The current study's findings give up new possibilities for clinical application, diagnosis, and treatment of ASD symptoms. Testing deficits in perceptual processing of acoustic voice parameters, such as vocal pitch, could be a simple supplementary tool in the diagnostic method of ASD, providing that our observations would be evaluated and boundaries for the distinction between ASD and non-ASD individuals would be quantified.

Furthermore, acoustic voice feature training could be used as a remedial method. There is evidence that basic auditory recognition, such as non-vocal pitch perception, can be enhanced through training in generally developed adults. Previous research suggests that non-vocal auditory perception training, such as musical training, increases higher-level auditory vocal perception, such as speech-in-noise perception (Slater. Speculatively, comparable training effects may be observed for the perception of vocal pitch, and vocal pitch training may improve higher-level auditory vocal processing, such as perception of vocal emotion or voice identification. However, it is still unknown if such training of vocal perceptual abilities will translate to performance in higher-level vocal processing. The majority of earlier research on speech perception in ASD focused on single tasks in one modality (for a review, see Baum et al. 2015) and within single voice perception components (see chapter 2.3). The current study's findings clearly indicate that this approach should be expanded to include testing cognitive and perceptual abilities in ASD. Assessing perceptive and higher cognitive abilities in the same group of people could help to explain the nature of the higher-level processing impairments that are part of the clinical diagnosis of ASD.

In the Sinhala speech synthesis component 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.

When considering the results obtained from the kids and toddlers regarding the development of a natural language processing component for ASD kids the following details can be displayed; With respective constructs, the best results produced 99.72 percent correct categorization for the entire sample. The best findings were 99.92 percent correct classification for male toddlers using 14 items and 99.79 percent correct classification for female toddlers using 18 items. Top results in boys produced 99.64 percent right classification with 18 items, whilst better outcomes in girls yielded 99.95 percent correct classification with 18 things. When employing 16 items, the best outcomes were 99.75 percent right classification when the education level was 15 years or fewer . When education level was 16 years or more , the results were virtually the same; that is, 99.70 percent correct classification was achieved using research instrument.

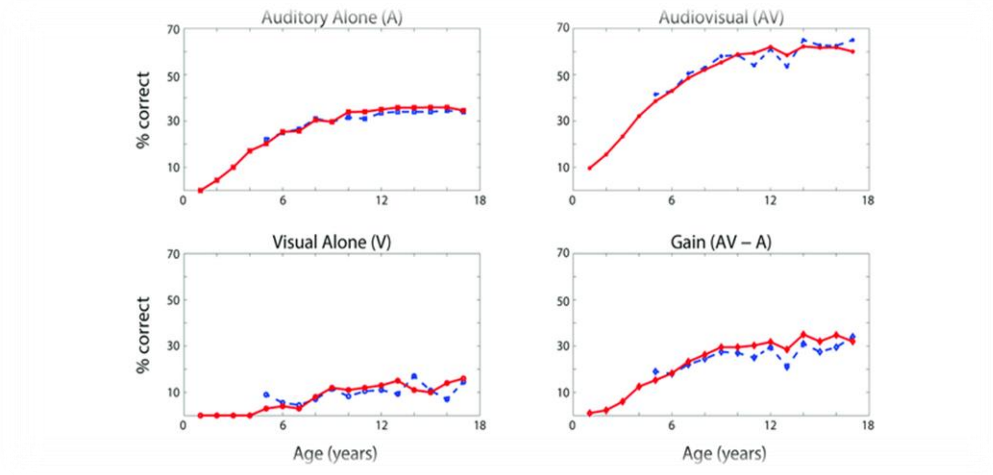
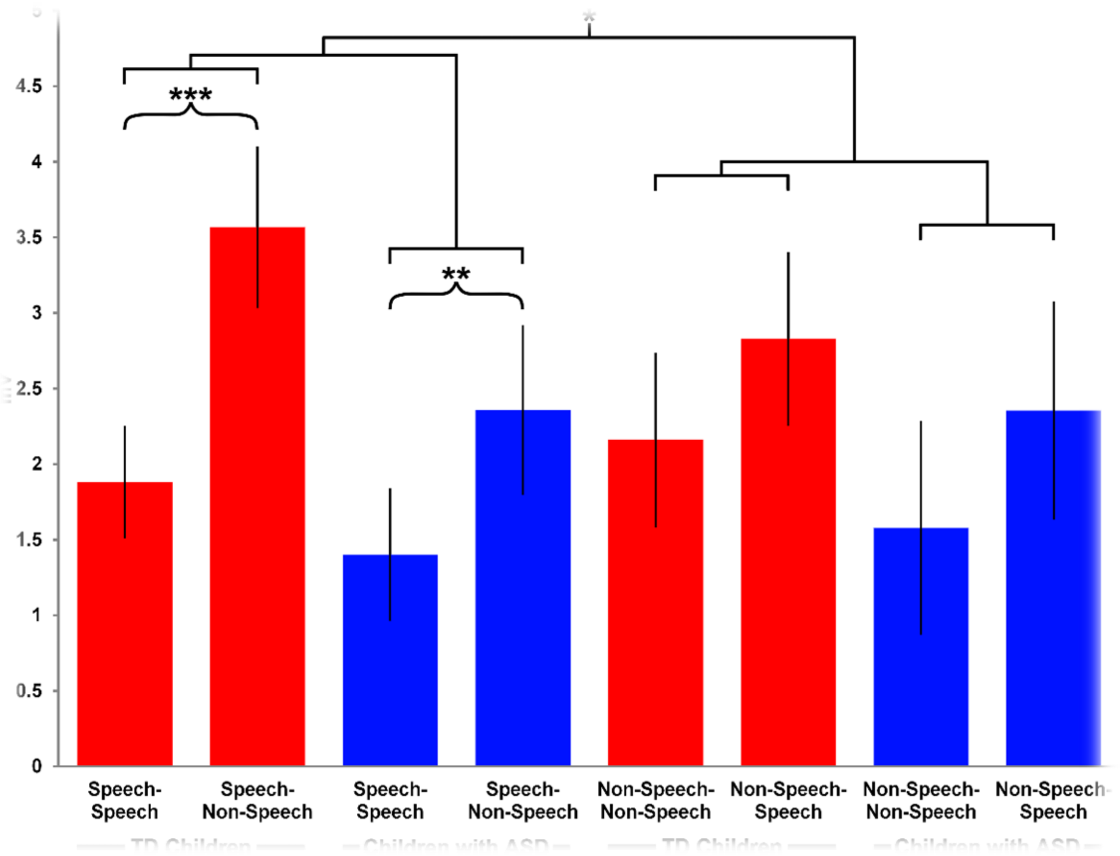


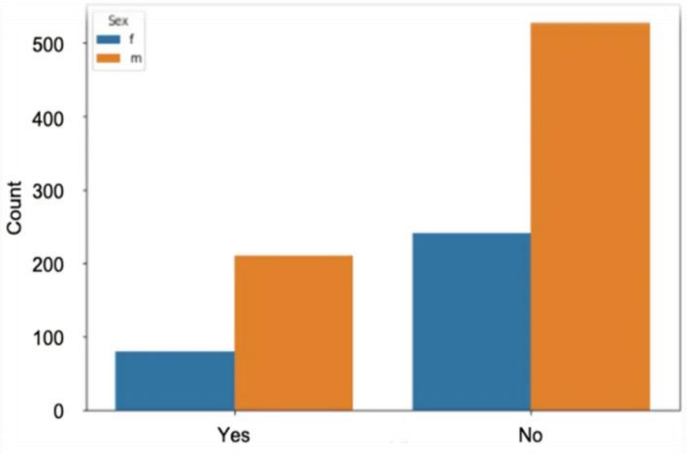
Figure 3.2: Speech recognition performance



Three consonant-vowel phrases of Sinhala, were recorded, digitized, and utilized to compute the SSG in the current study by a female English speaker. The glottal excitation waveform, formant frequencies for the three consonants in Sinhala, and formant frequencies for the vowel /a/ were examined. In order to achieve the same gross structure of stimuli, a 30 ms pre-voice bar was added to the stimuli, which is normally present in the /ba/ syllables. The consonant explosion lasted 10 milliseconds after the pre-constant voice bar. The consonant-to-vowel transition took 80 milliseconds, followed by a 60-millisecond steady-state vowel. The syllable and non-phonetic associated stimuli lasted 180 milliseconds in total. Five sinusoidal tones were used to construct the non-phonetic equivalents of the three speech syllables. The SSG calculated the frequencies and intensities of the tones, which were then picked based on the syllable format frequencies. The length and intensities of the non-phonetic stimuli, as well as the spectra of burst and burst-to-steady state format transitions, were preserved equal to those of the equivalent natural speech stimuli. As a consequence, synthetic stimuli differed only in terms of format transitions and plosives from matching voice stimuli. The tones' remaining acoustical characteristics, such as fundamental frequency, intonation, and intensity duration, were all the same. Despite the fact that speech and non-speech stimuli are physically identical, it has been demonstrated that both adults and children perceive speech stimuli as speech and non-speech stimuli as non-speech.

3.2 Research Findings

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.



We presented a conversational agent to assist teenagers with ASD in improving their communication abilities. We conducted an experimental study with five teenagers, analyzing their behavior over the course of several chats and interviewing them about their interactions with the system. According to the findings, the ASD.AI has the potential to benefit a large number of youngsters as a tool for developing communicative skills. Negative feedback, on the other hand, should be handled, and hypothetical queries should be avoided by the system.

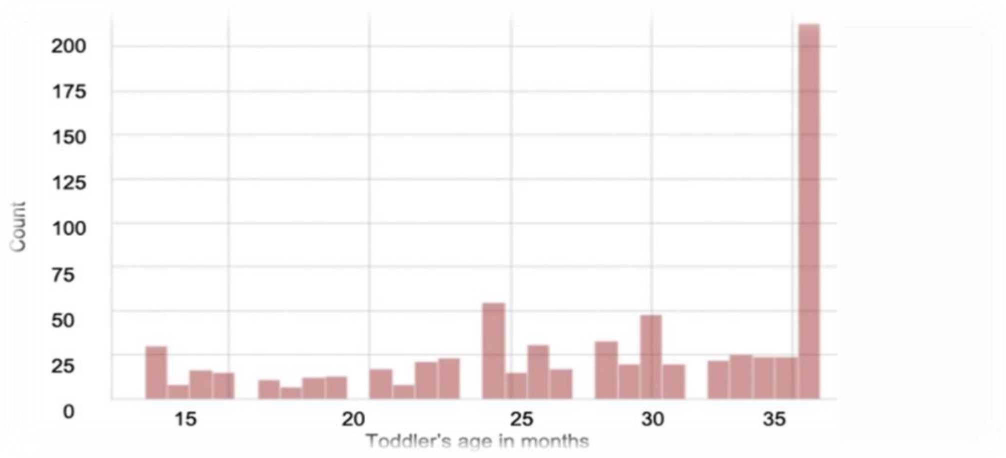
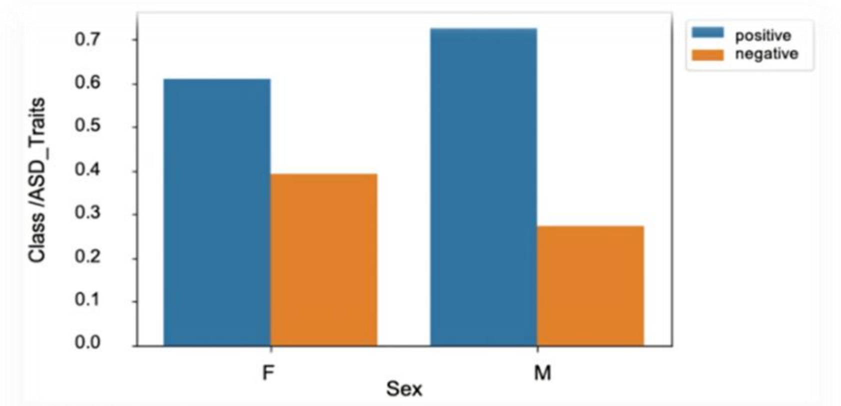


Figure 3.5: Age wise ASD detection in toddlers

The majority of ASD diagnostic cases in children occur around the age of 36 months. Between the ages of 15 and 20, the fewest cases were observed. Significant indications of autism appear at the age of three years, as shown in the graph (Fig. 3). Autism affects one in every 68 kids ages 2 to 3 years, according to the above figure.

The ASD features identified in male and female kids were placed on a gender distribution graph. As can be seen in the graph below, ASD is more prevalent in boys than in females.



Gender distribution of ASD traits

3.3 Discussions

The purpose of this study was to see how slightly elevated kids with ASD received vocal sounds while they were told to pay attentively to them. We discovered that ASD adults had noticeably shorter RTs for voices than it does for strings, similar to NT adults. Furthermore, while both experimental groups performed similarly on non-vocal sounds and tasks that did not require voice recognition, ASD people exhibited lower RTs for voice identification than the NT kids. The auditory chimeras give a new dimension to the comparison between groups. While the NT group did not retain the voice-processing benefit when temporally or spectral voice-specific knowledge was preserved, the ASD group appeared to do so, at least in part, as they had faster RTs for the chimeras when compared to non-vocal sounds. When examining our findings, it's vital to note that and between difference in RTs for vocal sounds cannot be explained by variations in motor planning, motor execution, task understanding, or other non-sensory abilities. The subjects in the ASD group compared with the control were IQ matched, and they performed identically in the simple RT task for all auditory stimuli, as well as in the go/no-go task for sounds of strings. As a result, the discrepancy can be attributed to the perceptual processing of tones of voice.

Different components of auditory processing were assessed using the simple RT task and the go/no-go task. The findings of the simple RT task demonstrate the ability to identify low-level auditory information quickly28. The acoustic characteristics of noises have been discovered to modify the basic RTs29 in certain research. The current findings, which reveal no significant differences in simple RTs between the four target types, suggest that the targets were acoustically well matched (in terms of onset time, for example) and equally easy to detect. Furthermore, both the ASD and NT groups performed equally well on the easy RT exercise.

For the go/no-go challenge, the findings were significantly different between the ASD and NT groups. Although earlier research has found that children with ASD have a lower preference for voices or name calling, we found that when ASD adults were taught to pay attention to auditory stimuli, their RTs for vocal sounds were faster than NT people. When ASD adults pay attention, their vocal processing speeds up even more. Furthermore, the ASD and NT groups had different types of noises that caused quick processing. Voices and auditory chimeras, particularly those possessing the temporal components of voices, were processed quickly by the ASD group. In two respects, these findings are startling. First, when confronted with vocal sounds, ASD children have been shown to have decreased neural activation in the voice-selective cortical area, which appears to give a neural basis for their lower sensory processing for voices. Second, in people with ASD, the processing efficiency of temporal modulated auditory stimuli is observed to be lower.

We can correlate our findings with the literature has provided the following innovative interpretation of ASD people's voice deficits detected without directed attention: Individuals with ASD have a level is required of vocal sounds that is intact, but they do not comprehend them holistically. The following are the proof or evidence from our findings: Participants in our study were given clear instructions to pay enough attention to the target noises, which included voices, and the Group exhibited outperformed the NT group. This shows that their representation of vocal sounds at a low level was preserved. At least few ERP investigations that contrast speech versus non-speech sounds have proposed preserved low-level speech representations. In addition, unlike the NT group, the ASD group reacted faster to auditory chimeras that incorporated both verbal and non-vocal cues. This could be due to the fact that they analyzed the sounds and relied on the many acoustic signals to a voice independently. While analytical attention may have helped them in our RT test with chimeras because they were unaware of the competing indications in the audio, it could be harmful when listening to genuine speech. It has been demonstrated that when ASD people listen to diverse sounds, their neural activity in the primary auditory cortex is higher than that of NT people, but their neural activity in the non-primary auditory system is lower, which supports this theory. This interpretation is in line with the improved visual performance theory, which describes how people with autism perceive local features more clearly. It also fits with the hypothesis of "weak central coherence"19. What remains to be determined would be whether the lack of impulsive orientation to voices and faulty brain connection between vocal processors and reward networks15 are a result of this lack of holistic processing, a cause of it, or if the two characteristics are co-occurring but separate. In conclusion, when ASD persons were directed to pay attention to auditory stimuli, their reaction time to voices in the go/no-go tasks was faster than when NT individuals were instructed to pay more attention to auditory stimuli. This finding suggests that the high-level deficits in vocal sounds seen in ASD are unlikely to be the result of impaired sensory representation. Rather, they could reflect problems with the overall processing of complex auditory clues to voices, as well as the voice's perceived reward value.

4. STUDENTS CONTRIBUTIONS

Table 1: Workload Distribution

Student Name	Student Registration Number	Description
Gunawardhana M.D.R.T	IT16090804	• Managed the ongoing conversation dialogues with the aid of a machine learning algorithm • Training based on previous conversation data using the machine learning algorithms. • To perform actions by interacting with an external system via API, query etc. • To add external system interaction to the conversation.
Anjali R.P.D.N	IT17109536	• Development of a way to distinguish a way between specific sound waves from the

		host according to the background noise. • If the same Sinhala word is pronounced in a different way development of a way that the software catches a correct way to recognize those voices. • Introduced an appropriate way to present the list of words that the system can identify.
Sampath G.A.D.M	IT16061880	• Developed a fully featured complete Sinhala Text to Speech system that gives a speech output similar to human voice while preserving the native prosodic characteristics in Sinhala language. • Developed a TTS system with the ability to maintain a real-time conversation with Autistic kids. • To develop a TTS system to pronounce the given text with proper rhythm, melody. • To find correct pronunciation, for different contexts in the text and to find correct intonation,

		stress, and duration from the text.
Herath H.M.D.N	IT18081794	• Developed a customized NLU tool with Sinhala language support as a component of Machine Learning based automated autism screening tool. Support both English and Sinhala languages • Developed an NLP chat-bot system to have conversations with kids to find their preferences. A prototype model is tested, and it will be converted to a chat-bot system to analyse each text received from speech recognition.

5. 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.

In recent years, machine learning and natural language processing (NLP) models have become hot topics in medicine, and they may be regarded a new paradigm in medical research. However, rather of creating wholly new knowledge, these procedures tend to support clinical theories, and only one significant segment of the population (i.e., social media users) is an imprecise cohort. Furthermore, several language-specific aspects can increase NLP approaches' effectiveness, and their extension to other languages should be researched further. Machine learning and natural language processing (NLP) approaches, on the other hand, generate important information from previously untapped data (e.g., patients' daily behaviors, which are normally inaccessible to care professionals). Before considering it as a supplement to mental health care, ethical concerns should be addressed as soon as possible. Machine learning and natural language processing methods may provide numerous viewpoints in mental health research, but they should also be viewed as clinical practice support tools. Over the last 25 years, the estimated prevalence of Autism has risen dramatically, and ASD diagnosis is time-consuming and labor-intensive. Our suggested ASD detection method has shown great promise in recognizing ASD based on the medical forms of the autistic kids. Our technology has the potential to drastically reduce the time spent waiting for an ASD diagnosis and help patients by permitting potential early intervention programs, which have been shown to be quite beneficial in many cases. Although the focus of this study is on ASD identification, the suggested NLP-based framework might be used to a variety of other inability conditions of kids such. In the future, we plan to develop a computerized index for ASD patients that will indicate their severity based on their medical data and may be used to track their improvement over time. Changes in the index could also be used as an outcome metric in clinical trials of various medicines.

REFERENCES

[1] "Do-it-yourself NLP for Bot developers – Rasa Blog – Medium", Medium, 2016. [Online]. Available: https://medium.com/rasa-blog/do-it-yourself-nlp-for-bot-developers2e2da2817f3d. [Accessed: 09- Sep- 2018].

[2] Hettige, B. and Karunananda, A. (2006). First Sinhala Voice assistant in action. [online] Staffweb.sjp.ac.lk. Available at: http://staffweb.sjp.ac.lk/sites/default/files/budditha/files/budditha2006.pdf [Accessed 23 May 2018].

[3] Deep Learning Based Voice Assistant Models. [online] Research Gate. Available at: https://www.researchgate.net/publication/323587007_Deep_Learning_Based_Voice assistant_Mode ls [Accessed 21 Jul. 2018].

[4]"iOS - Siri", Apple. [Online]. Available: https://www.apple.com/ios/siri/. [Accessed: 10- Sep- 2018].

[5]"Rasa: Open-source conversational AI", Rasa.com. [Online]. Available: https://rasa.com/ [Accessed: 11- Sep- 2018].

[6]"Snips Natural Language Understanding — Snips NLU 0.16.5 documentation", Snipsnlu.readthedocs.io. [Online]. Available: https://snips-nlu.readthedocs.io/en/latest/. [Accessed: 11- Sep- 2018].

[7]"An Overview of Voice assistant – Voice assistants Life", Voice assistants Life. [Online]. Available: https://voice assistantslife.com/an-overview-of-voice assistant-a539b5fc55d3. [Accessed: 11- Sep- 2018].

[8] Dialogflow Entities: Identify things your users mention [Basics 2/3]. Google, 2018.

[9]"Dialogflow", Dialogflow, 2018. [Online]. Available: https://dialogflow.com/. [Accessed: 14- Sep- 2018].

[10] "LUIS: Language Understanding Intelligent Service", Luis.ai, 2018. [Online]. Available: https://www.luis.ai/home. [Accessed: 14- Sep- 2018].

[11] Cho, K., Van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation.

[12] Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems, pages 3104–3112.

[13] "Voice assistant Market Survey- 2017: Mindbowser Info Solutions", Mindbowser, 2018. [Online]. Available: http://mindbowser.com/voice assistant-market-survey-2017/. [Accessed: 10- Aug- 2018].

[14] M. KOTTORP and F. JÄDERBERG, Voice assistant as a potential tool for businesses. KTH SKOLAN FÖR INDUSTRIELL TEKNIK OCH MANAGEMENT, 2017, pp. 4-18.

[15]"NLUlite", Nlulite.com, 2019. [Online]. Available: http://nlulite.com/. [Accessed: 04- Apr- 2019].

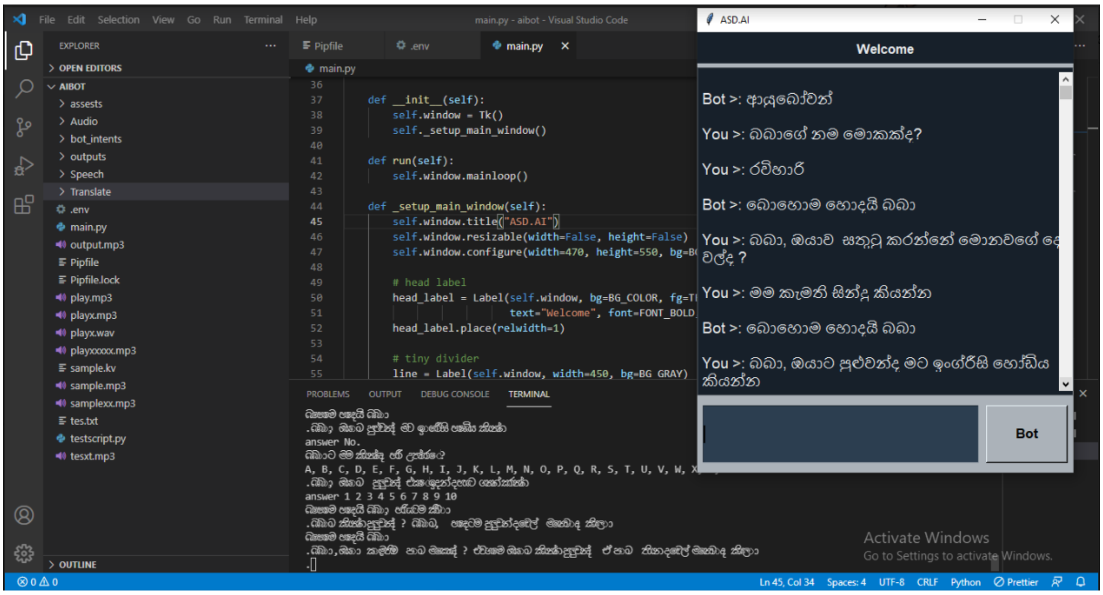
[16] "Supervised Word Vectors from Scratch in Rasa NLU", Medium, 2019. [Online]. Available: https://medium.com/rasa-blog/supervised-word-vectors-from-scratch-in-rasa-nlu-6daf794efcd8. [Accessed: 24- Apr- 2019].

[17]"tensorflow/models", GitHub, 2019. [Online]. Available: https://github.com/tensorflow/models/tree/master/research/syntaxnet. [Accessed: 24- Apr- 2019].

[18] "Rasa Stack: Open-source conversational AI", Rasa.com, 2019. [Online]. Available: https://rasa.com/products/rasa-stack. [Accessed: 11- May- 2019].

APPENDICES

Appendix 1 - Implementation

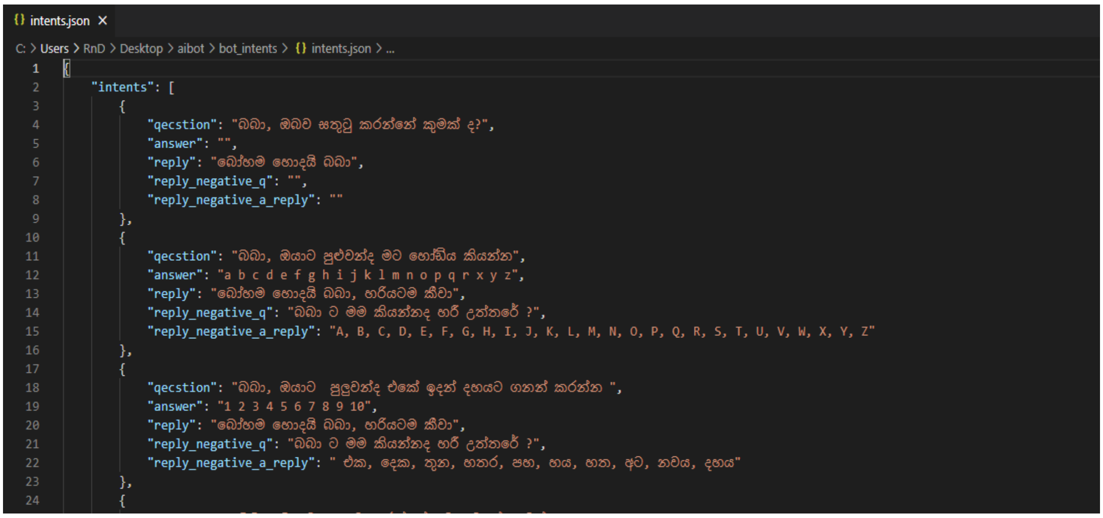


```python
    def __init__(self):
        self.window = Tk()
        self._setup_main_window()

    def run(self):
        self.window.mainloop()

    def _setup_main_window(self):
        self.window.title("ASD.AI")
        self.window.resizable(width=False, height=False)
        self.window.configure(width=470, height=550, bg=B

        # head label
        head_label = Label(self.window, bg=BG_COLOR, fg=T
                           text="Welcome", font=FONT_BOLD
        head_label.place(relwidth=1)

        # tiny divider
        line = Label(self.window, width=450, bg=BG GRAY)
```

```
බොහොම හොදයි බබා
.බබා, ඔයාට පුළුවන්ද මට ඉංග්‍රීසි හෝඩිය කියන්න
answer No.
බබාට මම කියන්නද හරි උත්තරේ?
A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X
.බබා, ඔයාට පුළුවන්ද එකේ ඉදන් දහයට ගනන් කරන්න
answer 1 2 3 4 5 6 7 8 9 10
බොහොම හොදයි බබා, හරියටම කීවා
.බබාට කියන්න පුළුවන්ද ? බබාට, [illegible]
බොහොම හොදයි බබා
.බබා, [illegible] ? [illegible]
.
```

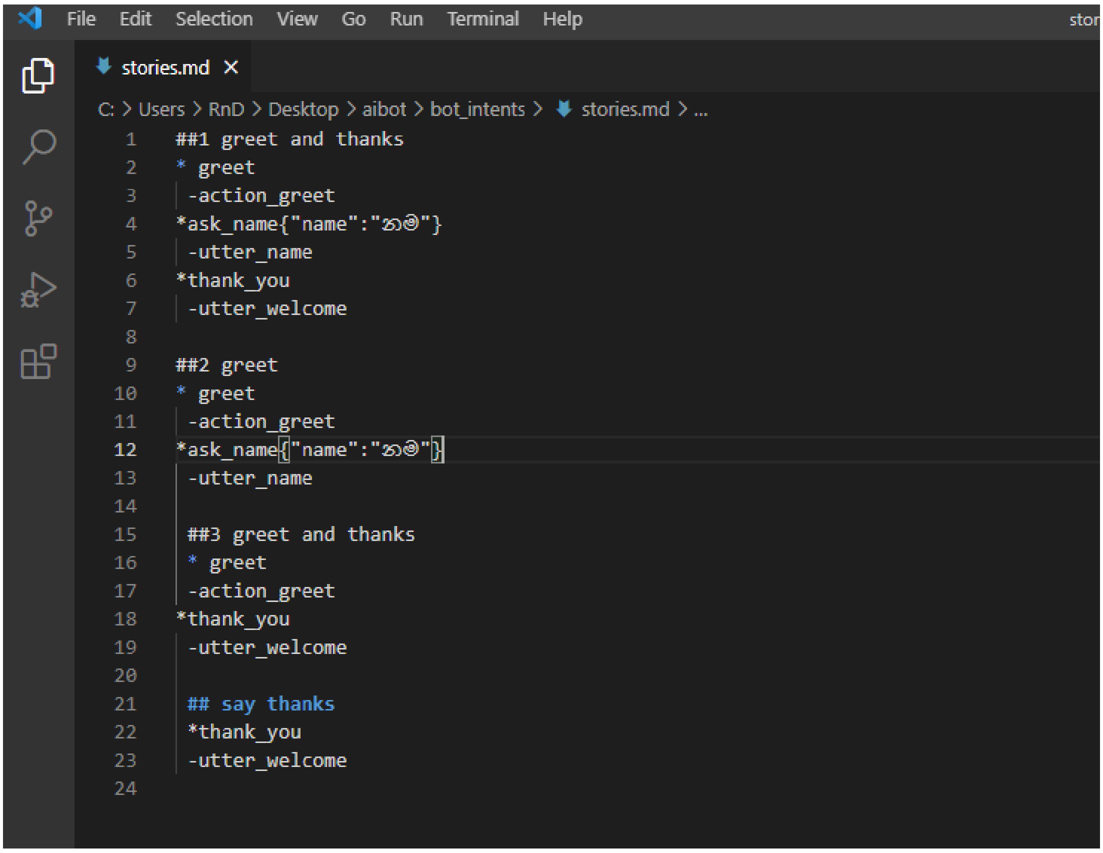
```
Bot >: ආයුබෝවන්

You >: බබාගේ නම මොකක්ද?

You >: රවිහාරි

Bot >: බොහොම හොදයි බබා

You >: බබා, ඔයාව  සතුටු කරන්නේ මොනවගේ දෙ
වල්ද ?

You >: මම කැමති සින්දු කියන්න

Bot >: බොහොම හොදයි බබා

You >: බබා, ඔයාට පුළුවන්ද මට ඉංග්‍රීසි හෝඩිය
කියන්න
```

Appendix 2 - Intents File (intents.json)


intents.json
C: > Users > RnD > Desktop > aibot > bot_intents > {} intents.json > ...

```json
{
    "intents": [
        {
            "qecstion": "බබා, ඔබව සතුටු කරන්නේ කුමක් ද?",
            "answer": "",
            "reply": "බෝහම හොදයි බබා",
            "reply_negative_q": "",
            "reply_negative_a_reply": ""
        },
        {
            "qecstion": "බබා, ඔයාට පුළුවන්ද මට හෝඩිය කියන්න",
            "answer": "a b c d e f g h i j k l m n o p q r x y z",
            "reply": "බෝහම හොදයි බබා, හරියටම කීවා",
            "reply_negative_q": "බබා ට මම කියන්නද හරි උත්තරේ ?",
            "reply_negative_a_reply": "A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z"
        },
        {
            "qecstion": "බබා, ඔයාට  පුලුවන්ද එකේ ඉදන් දහයට ගනන් කරන්න ",
            "answer": "1 2 3 4 5 6 7 8 9 10",
            "reply": "බෝහම හොදයි බබා, හරියටම කීවා",
            "reply_negative_q": "බබා ට මම කියන්නද හරි උත්තරේ ?",
            "reply_negative_a_reply": " එක, දෙක, තුන, හතර, පහ, හය, හත, අට, නවය, දහය"
        },
        {
```

Appendix 3: Stories File (stories.md)


File
Edit
Selection
View
Go
Run
Terminal
Help
stor
stories.md ×
C: > Users > RnD > Desktop > aibot > bot_intents > stories.md > ...

```
##1 greet and thanks
* greet
 -action_greet
*ask_name{"name":"නම"}
 -utter_name
*thank_you
 -utter_welcome

##2 greet
* greet
 -action_greet
*ask_name{"name":"නම"}
 -utter_name

 ##3 greet and thanks
 * greet
 -action_greet
*thank_you
 -utter_welcome

 ## say thanks
 *thank_you
 -utter_welcome

```