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

Project Proposal Report

(Proposal documentation submitted in partial fulfillment of the requirement for the Degree of Bachelor of Science Special (honors) In Information Technology)

Bachelor of Science Special (Hons) in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

January 2021

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

Project Proposal Report

Student ID	Name of the student
IT16090804	Gunawardhana M.D.R.T.
IT18081794	Herath H.M.D.N.
IT17109536	Anjali R.P.D.N.
IT16061880	Sampath G.A.D.M.

Supervisor: Prof. Koliya Pulasinghe

Degree of Bachelor of Science Special (honors) In Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

January 2021

Declaration of the Candidate and Supervisor

We declare that this is our own work, and this proposal does not incorporate without acknowledgement of any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Student ID	Name of the student	Signature
IT16090804	Gunawardhana M.D.R.T.	
IT18081794	Herath H.M.D.N.	
IT17109536	Anjali R.P.D.N.	
IT16061880	Sampath G.A.D.M.	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor: Date:

Signature of the Co-supervisor: Date:

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. Koiya Pulasinghe for providing insight and expertise that greatly assisted the research and for supervising our project.

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.

Abstract

Autism spectrum disorder is an umbrella term for a group of neurodevelopmental disorders that are associated with impairments to social interaction, communication, and behavior. It is characterized by impairments in social-communication and the manifestation of repetitive or unusual behaviors. Recently, there has been a sharp increase in the prevalence of ASD. For example, data from the Center for the Disease Control (CDC) in the United States suggests 1 in 68 children are identified as having a diagnosis of ASD. Over the past decade, many pediatricians have begun screening for autism. The purpose of an autism screening is to identify common early signs of autism. All children start developing language from the day they are born. It can be harder for autistic children to learn and use language than typically developing children. Autism awareness is growing in Sri Lanka, but it is growing slowly and only through the hard work of institutions, initiatives, and dedicated individuals.

Typically, autism spectrum disorder is first detected with a screening tool (e.g., modified checklist for autism in kids). However, the interpretation of autism spectrum disorder behavioral symptoms varies across cultures: the sensitivity of modified checklist for autism in kids is as low as 25 percent in Sri Lanka. So, a culturally sensitive screening tool for autism assessment has become a major requirement. Low- and middle-income countries have a shortage of mental health specialists, which is a key barrier to obtaining an early autism spectrum disorder diagnosis. Early identification of autism spectrum disorder enables intervention before atypical patterns of behavior and brain function become established. This article proposes a culturally sensitive autism spectrum disorder screening tool. The proposed application embeds an intelligent machine learning model and uses a clinically validated symptom checklist to monitor and detect autism spectrum disorder in low- and middle-income countries like Sri Lanka for the first time.

Table of Contents

Declaration of the Candidate and Supervisor iii

Acknowledgment iv

Abstract v

List of Figures vii

List of Tables vii

List of Abbreviations vii

1. Introduction 1

1.1. Background and Literature Review 1

1.1.1. Dialogue Management 3

1.1.2. Natural Language Processing 4

1.1.3. Speech recognition 4

1.1.4. Speech Synthesis 5

1.2. Research Gap 6

1.3. Research Problem 7

2. Objectives 9

2.1. Main Objective 9

2.2. Specific Objectives 9

3. Methodology 10

3.1 High Level Architecture Diagram 12

3.2 System Architecture Diagram 13

4. Testing & Evaluation 14

5. Description of Personal and Facilities 15

5.1. Work Breakdown Structure 15

5.2. Gantt Chart 16

5.3. Workload Allocation 17

6. Conclusion 19

References 20

List of Figures

Figure 3.1.1 12
Figure 3.2.1 13
Figure 5.1.1 15
Figure 5.2.1 16

List of Tables

Table 5.2.1 18

List of Abbreviations

Abbreviation	Description
RNN	Recurrent Neural Network
GPU	Graphical Processing Unit
API	Application Programming Interface
NLU	Natural Language Understanding
STT	Speech to Text
VB	Variational Bayesian
DSR	Distributed Speech Recognition
TTS	Text to Speech
LMIC	Low to Middle Income Countries
ASD	Autism Spectrum Disorder
PAAS	Platform as a Service

1. Introduction

1.1. Background and Literature Review

The advancement of technology allows various approaches of formal and informal autism screening tools. These can range from simple observations to formal assessments. Some of the more commonly used autism screening tools are:

- Modified Checklist for Autism in Toddlers, revised (M-CHAT), is a popular 20-question test designed for toddlers between 16 and 30 months old. Note: there is recent evidence that the M-CHAT may not be as effective in screening females and minority, urban and low-income children.
- The Ages and Stages Questionnaire (ASQ) is a general developmental screening tool that examines developmental challenges at specific ages.
- Screening Tool for Autism in Toddlers and Young Children (STAT) is an interactive screening tool comprising of twelve activities that assess play, communication, and imitation.
- Parents' Evaluation of Developmental Status (PEDS) is a general developmental parent interview designed to identify delays in motor, language, self-help, and more.

Autistic children can find it hard to relate to and communicate with other people. They might be slower to develop language, have no language at all, or have significant problems with understanding or using spoken language. They might not use gestures to make up for the problems they have with words. Autistic children tend to communicate mostly to ask for something or to protest. They are less likely to communicate for social reasons, like sharing information.

They also often have difficulty knowing when and how to communicate with people in socially appropriate ways. For example, they might not make eye contact or let another person take a turn in a conversation.

To communicate effectively, children need to understand what other people say to them (receptive language) express themselves using words and gestures (expressive language) use their receptive and expressive language skills in socially appropriate ways.

Autism spectrum disorder (ASD) is a developmental disorder that affects social interaction, communication, and behavior. ASD can be diagnosed at any age, but symptoms manifest within 24 months of birth. However, less evidence is available for ASD prevalence estimates in low- and middle-income countries (LMIC) like Sri Lanka. ASD detection is poor in LMIC compared with developed countries due to research and funding limitations. To date, Sri Lanka only completed limited research to determine how many of its citizens are autistic, and health officials who say ASD is non-existent in this region likely do not know how to identify it. This is a deeply concerning issue, and there is an urgent need for more support and services for these individuals living in this country. Early identification of ASD in children enables intensive intervention before neuronal pruning is completed.

By considering these facts, this research proposal has proposed an autism screening tool consist of intelligent artificial agents known as "Voice Bots" to automate the speaking process with the ability to handle operations without human intervention to identify the children with autism by their voice.

The advancement of artificial intelligence chatbot with traditional text-based interface came to the market with it being the new phenomenon in the most recent years. But the main flow associated with chatbots is the lack of human emotion. Lack of human emotion causes chatbots to seem extremely unnatural and semi-problematic. As a solution to this problem voice-activated chatbots and products came into the market with Amazon Alexa, Google Assistant, Apple Siri becoming insanely popular as virtual assistants. Voice-enabled chatbot frameworks are mainly composed of the following research components.

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

Researchers have conducted numerous researches around the above areas to achieve the best result possible in each area of application. Following were some significant findings we have encountered in the above domains.

1.1.1. Dialogue Management

Historically dialogue management systems have been built based on a custom pipeline with separate modules for language understanding, state tracking, action selection, and language generation combined to provide the result. With the above approach, each module must be trained independently with labeled data. But considerable complexity in the engineering process and tightly coupled dependencies between modules urged researchers to find alternative solutions. Recent advancements of recurrent neural networks allowed researchers to implement solutions that infer a latent representation of the state, but it lacked a general method to inject domain knowledge and constraints. As state by [1] training a recurrent neural network on text transcripts of dialogue can have the RNN infer a latent representation of state eliminates the need for state labels. Furthermore, [1] the proposes a solution with the developer having the ability to express the domain knowledge for the scenario using via software and action templates rather than allowing dialogue management system to learn it from the dialogue. This approach supports the separation of concerns where domain knowledge and constraints can be expressed in software with control flow is learned from these inputs.

It affords more developer control and requires a limited amount of data to train the system initially. According to the results posted by the above research, it outperforms the performance of purely learned models and rule-based systems. It is also suggested that the integration of reinforcement and supervised learning in the current solution can further optimize the system.

According to the [2] current neural model-based dialogue generation seems like a proven method while the only downside is completely ignoring the future outcome of the dialogue in the dialogue generation process. Reinforcement learning algorithms were applied to the above issue so that generated dialogues will consider the future outcome of the response via a predefined reward function. As discussed by the above research paper application of deep reinforcement learning researchers were able to optimize reward based interactive responses that foster a more sustained conversation. [3] The inclusion of a transfer learning approach to an existing goal-oriented dialogue system within a closed domain can speed up the learning rate of the system by 5 to 10 times with more than 20% success rate in response generation. Compared to a deep reinforcement learning approach this implementation can be used within a space of low volume of data. This method also improves the success rate of the system significantly even with the availability of high-volume domain-specific data. Most of the research were significantly divided among dialogue management using flat reinforcement learning agents and rule-based agents. The latter approach depends on a deterministic set of rules through a hand-coded approach and lacks a high-level description of a dialogue system [4]. Formulation of complex tasks [5] in a mathematical framework of options over Markov Decision Processes (MDPs), and proposing a hierarchical deep reinforcement learning approach to learning a dialogue manager that operates at different temporal scales can significantly outperform a flat reinforcement learning agent and rule-based agents. This research introduces a dialogue management engine that consists of three components, a top-level sub-task selector that selects a subtask or an option for the given input, a low-level dialogue policy that selects primitive actions for the above-selected sub-task, finally a global state tracker object that spans across the entire conversation tree to maintain the future outcome of the dialogue. According to the test results of the research hierarchical structure of the agent improved the coherence of the dialogue flow.

1.1.2. Natural Language Processing

The research work in natural language processing has been increasingly addressed in recent years. Natural language processing is the computerized approach to analyzing text and being a very active area of research and development. [6] Once a Research has proposed a platform with deep learning methods integrated with GPU based training on the mind. The research focuses mainly on core semantic problems, likely including efforts to generalize semantic role labeling to all words, models for general coreference resolution, semantic parsers that build relatively compete for meaning representations, and approaches for semi-supervised learning of improved word representations. In contrast to the above research [7] presents a modular natural language processing(Rasa NLU) decoupled from the dialog management unit. Its API is based on sckit-learn and Keras to focus on consistent API over a strict inheritance. Rasa NLU uses a fastText approach for text classification with pre-trained word embeddings with trained intent classifiers. The comparison [8] of its performance with alternative solutions shows Rasa NLU taking the upper hand in benchmarks.

1.1.3. Speech recognition

The main task is to get a computer to understand oral language. By “understand” it means to behave appropriately and convert the input speech into another medium, in this case, text. Speech recognition is therefore referred to as speech-to-text (STT). Even though this topic seems to be probed by many researchers now, it has existed since 1920 when machine recognition was introduced. Since that engineers and scientists have been working on several methods/patterns which became advanced from time to time. Some of these are,

1. Acoustic phonetic based speech recognition
2. Hardware-based recognizer
3. Pattern-based speech recognition
4. Continuous word-based speech recognition
5. Hybrid statistical and connectionist (HMM/ANN) based speech recognition.
6. Variational Bayesian (VB) estimation-based speech recognition

Some of the ongoing research can be described as below.

According to [9] The Aurora framework, which is developing standards for Distributed Speech Recognition (DSR) where the speech analysis is done in the telecommunication terminal and the recognition at a central location in the telecom network. The framework is currently being used to evaluate alternative proposals for front-end feature extraction [3]. Furthermore, [10] an empirical comparison among the CTC, RNN-Transducer, and attention-based Seq2Seq models for end-to-end speech recognition has been implied. The outcome would be to emphasize that without any language model, Seq2Seq and RNN-Transducer models both outperform the best reported CTC models with a language model, on the popular Hub5'00 benchmark.

1.1.4. Speech Synthesis

Each spoken word is created from the phonetic combination of a set of vowel and consonant speech sound units. The process of converting any arbitrary text in any given language into a corresponding speech sound unit is known as Speech Synthesis [11], [12]. This is also called as Text-To-Speech (TTS). Lots of researchers choose this topic for their research as it is currently a trending topic in the information technology field. But this topic is in this field since the 18th century [13]. But with the machine learning approach, this gained more attraction from the researchers in this field. From the beginning of this field, many types of research have been done based on many languages. Such as Slovak [14], Indian languages – Tamil, Hindi, Malayalam, and Telugu [15], Indian languages of Devanagari script [16], Indonesian [17], Moroccan Arabic [18], etc.

These researches have done using various methods and techniques based on the purpose of their research and the language they focused on. But the main purpose remains the same which is to convert a text format to a sound signal. There are many speech synthesis methods including, [12]

1) Unit Selection Synthesis: This method uses a large database of recorded words and because of this database it provides more naturalness to the output.
2) Diaphone Synthesis: This method uses a relatively small database than unit selection and it store small units of speech. Because of this output is also less naturalness than output of unit selection synthesis.
3) Domain Specific Synthesis: This method is mostly used in systems which need small amount of vocabulary.
4) Formant Synthesis: This method is based on the source-filter model. And there are two types of structures namely Cascade and Parallel. Combination of these two types can also be used get more performance.
5) Articulatory Synthesis: This is based on modelling of the human speech production system and this is hard to implement.
6) Hidden Markov Model [12],[19]

Since the input is unpredictable when using speech synthesis, to overcome the unpredictability of the input researchers moved towards a context-independent approach. Even humans can pronounce new words efficiently from the experiences of other words. This is the base of this context-independent approach.

1.2.Research Gap

A significant amount of research has been carried out in the following research domains like Speech Recognition, Conversational Artificial Intelligence, and Speech Synthesis which are related to the proposed solution discussed in this research proposal. Many of the research conducted has enabled the following tasks to be executed from the products which are like the proposed solution in this paper.

The Advancement of Speech Recognition technologies allowed platforms to support real-time speech recognition with a wide array of different language support. According to the research carried out on each platform most platforms do not support the speech recognition process. With currently available speech synthesis platforms users are limited to the voices provided by the platform which makes most of them the automated presence to the listener. It was difficult to find a platform with support for training custom voices to provide more naturalistic voice output after the process. Even though most of these platforms support languages across the world we were yet to find a single platform with support for the Sinhala language.

Related to Dialogue Management different approaches have been taken by researchers to achieve the end goal of task completion with the majority focusing on the English Language. Some research focuses on creating an initial knowledge base to map conversation based on entities and actions. In those systems, the output will be generated by mapping the input source content into knowledge base entities and actions.

These types of research have been improved by adding the ability to manage the state of the current conversation thus converting old stateless dialogue management into stateful dialogue management. The advancement of GPU processing and machine learning allowed dialogue management research to focus on a deep learning-based neural network approach. Recent research in this field were mostly carried out on this approach. Sinhala Language support issue was still prominent in these research areas too.

But the proposed platform will mainly focus on providing Sinhala Language support in Speech Recognition, Conversational Artificial Intelligence, and Speech Synthesis. The platform will be a standalone solution that can be installed locally with the ability to customize according to the business scenario. Also, the machine learning algorithms will be trained using existing data to create a domain-specific representation rather than a generic approach followed by most other platforms.

1.3.Research Problem

Day by day the technology becomes better and better making the life of people easier. But there are some sides which are left untouched. Due to cultural reasons, ASD awareness is low in LMIC like Sri Lanka. Resource constraints meant that when ASD is identified, patients are often left untreated for long periods of time. Early identification and diagnosis are important to improve clinical outcomes of small children with ASD. Smart devices represent an ideal platform for a computer-aided tool, as they are highly accessible and prevalent across the world. Most existing screening tools automate standard screening checklists such as M-CHAT-

R/F. Only the ASD.AI application embeds an intelligent machine learning model to arrive at a decision. This article proposes a novel tool for ASD screening using a culturally sensitive symptom checklist and embedded machine learning model. A variety of supervised learning models were trained on PAAS data collected clinically. The proposed application has shown greater predictive performance than current paper-based methods (PAAS). The new application is important to improve ASD awareness and detection, by enabling non-specialist healthcare workers to screen for ASD during home visits. Furthermore, valuable data can be collected about the prevalence of ASD in LMIC (which is currently scarce) and resources allocated correctly to decrease treatment delays.

2. Objectives

2.1. Main Objective

The primary objective of the proposed solution is to create a machine learning based automated autism screening tool to reduce or eliminate error-prone, inefficient human intervention in the field. The proposed system's ability to support both English and Sinhala languages. The system with its efficient and robust performance will have a direct implication on the quality of the service.

2.2.Specific Objectives

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

- Increase availability.

The intelligent agents deployed by the proposed platform can be operational and readily engage with their defined goal 24 hours a day, 365 days a year. As humans are emotional being their current mental status can have a direct implication on the quality of service provided by the human. But with the intelligent agent's probability of this incident occurring is never.

- Handle many requests at a time with the proposed tool.

With the proposed platform intelligent agents will be deployed depending on the current load of requests to be handled and the agents will have the ability to handle multiple conversations simultaneously compared to its human counterpart.

- Decrease cost.

The system is easy to configure to meet different needs from time to time. Once deployed, intelligent agents will ultimately have only little to known maintenance cost compared to current systems. The system will allow easy adaptability across different languages because of the modular platform it was built upon.

- Increase overall productivity of the service.

Intelligent agents once deployed by the proposed platform will be reactive and efficient. Due to the intelligent agent's ability to interact with multiple users at a time, users of the system can gain in productivity, time, and scalability.

3. Methodology

The proposed solution will be based on the existing research found in the key areas of the platform such as Speech Recognition, Conversational Artificial Intelligence, and Speech Synthesis with the unique capability of Sinhala Language support and a being standalone decoupled solution. The solution will heavily depend on machine-learning algorithms to achieve human-level intelligence in decision making, dialogue generation and to take actions that maximize its chance of successfully achieving the end goal of screening the kids with autism.

Initially, the kids' voices will be input into the system. The system's noise reducer removes the noise attached to the current voice file. Then speech recognition engine developed for the proposed solution based on the deep speech implementation [20] proposed by researchers at Baidu, further following the most recent version of implementation being version 3 [10] extracts entities and actions associated with the current voice data into text format. The provided implementation will be tweaked accordingly to support Sinhala speech recognition. The key factor for selecting the above implementation relies on the architecture proposed by the above implementation. Rather than traditional text-to-speech engines with laboriously engineered processing pipelines and poor performance in noisy environments the selected approach is a well-optimized RNN training system that uses multiple GPUs to train on a large amount of varied data. This gives us the ability to train Sinhala language-specific data without having to implement custom pipelines and phoneme dictionary for the language. Further, the implementation highlights its ability to handle challenging noisy environments in the proposed system. The selected solution even supports training via a labeled data set of transcripts. The machine learning model for speech recognition will be trained from the corpus of existing Sinhala conversations for the Sinhala language. Furthermore, for the English language, Mozilla's common voice data set will be used.

The natural language processing capability of the proposed platform will extend the open-source implementation provided by the Rasa framework [7] by adding the support of Sinhala language processing. After receiving the voice data as text via speech recognition the natural language processing engine will extract the intent, entities, and any other structured information from the provided text in the Sinhala language. The processed data from the natural language processing unit will be passed to the dialogue management unit[7]. The dialogue management unit which is responsible for generating responses for user queries with the end goal of task completion will be developed considering different implementations proposed by the recent research for high accuracy and speed. The dialogue management engine architecture will closely follow the architecture proposed by the Rasa dialog management unit. The main purpose of using the above architecture is its ability to support machine correction where users can respond whether the action predicted by the engine is correct or wrong depending on the scenario. These functionalities allow the platform to suggest actions to human agents in the initial pace of the platform to evaluate itself. Additionally, the platform can visualize a graph of training dialogues which can be then be converted as a knowledge base for the business domain. The dialog management unit proposed [7] will be modified by applying deep reinforcement learning [2] with the ability to generate utterances that optimize future re- ward, successfully capturing global properties of a good conversation. This modification allows more diverse, interactive responses that foster a more sustained conversation.

The proposed solution will then convert the response provided by the dialogue management unit to Voice via speech synthesis to mimic a human voice to generate a human-like conversation between kids. The speech synthesis engine should support real-time Sinhala text to speech conversion with the absolute minimum amount of lag in the process to keep the conversation between customer and bot more realistic. To achieve the above-described efficiency and performance speech synthesis engine will be built based on the Deep Voice 3 [10] implementation proposed by researchers at Baidu. The above architecture supports 116 queries per second on a single GPU server which can support the above use case of real-time text to speech conversation with minimum lag or no lag at all. This architecture is capable of multi-speaker speech synthesis to distinguish each voice conversation to eliminate robotic response generation which is present in currently available commercial systems.

3.1 High Level Architecture Diagram

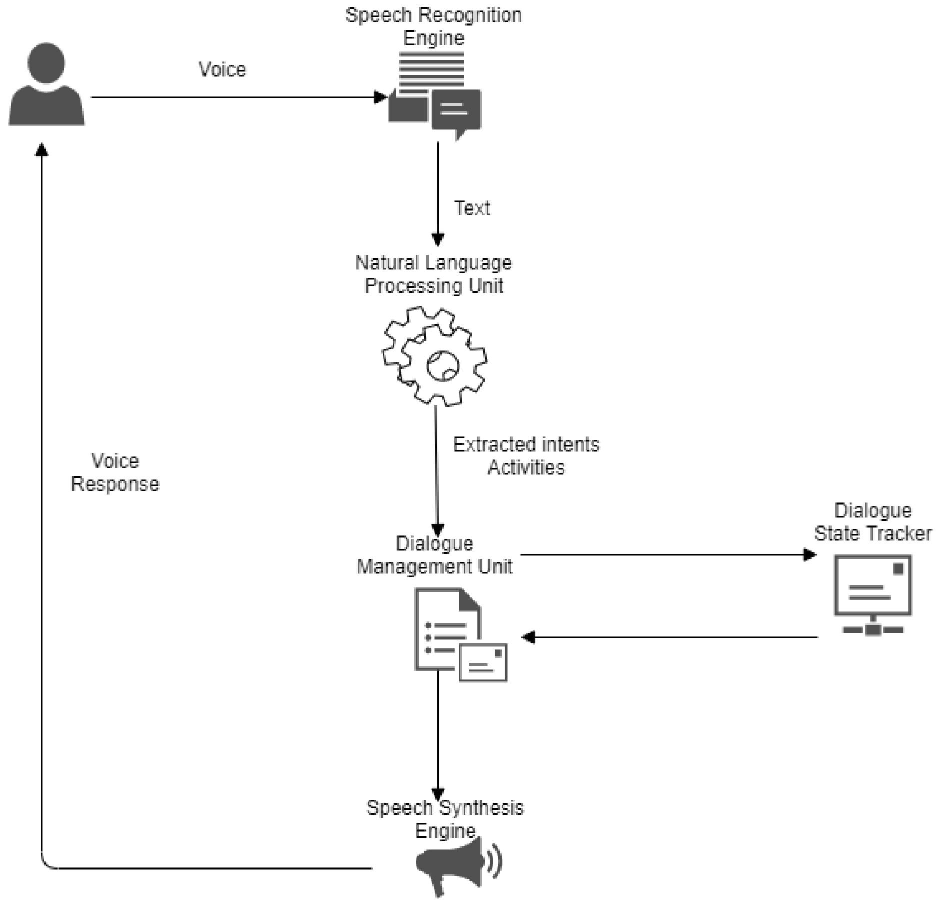

Figure 3.1.1: High Level Architecture

3.2 System Architecture Diagram

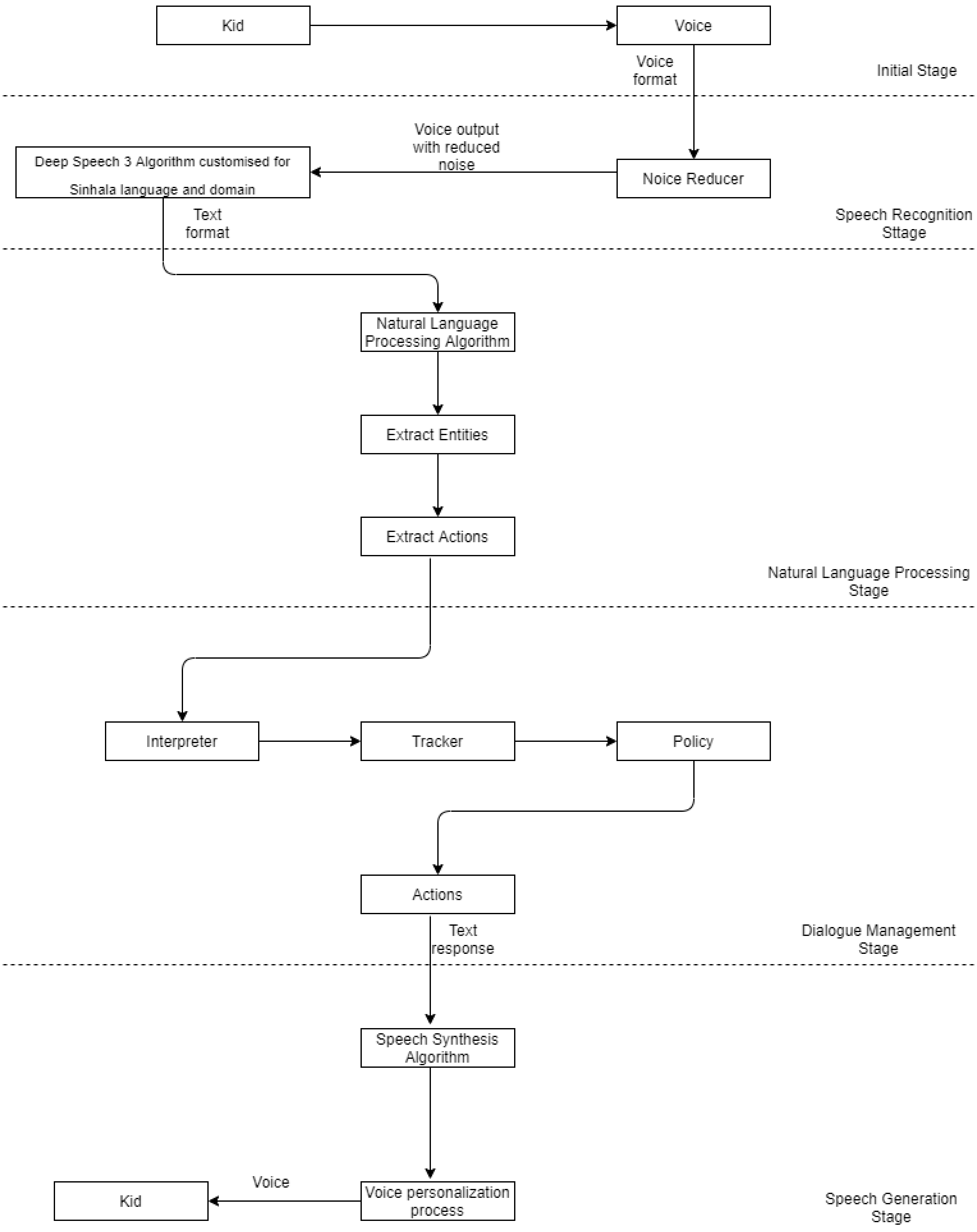

Figure 3.2.1: System Architecture

4. Testing & Evaluation

As the system depends heavily on deep learning algorithms it requires an abundance of label data for speech recognition, natural language processing, and speech synthesis. For the proposed system we need voice conversations with its transcript in-order to train machine learning models. To achieve the above requirement voice conversation data with its generated transcript will be used provided by an existing customer service provider for Sinhala language consist of nearly 1000 hours of data. Regarding English language-based training public datasets like Mozilla common voice will be used. The initial testing process will be carried with the assistance of human agents. At the start system will suggest actions to the human agent based on the ongoing conversation with the kid. The human agent will mark each action as correct or incorrect. Based on the responses by the human agent the system will tweak its responses accordingly and a performance score will be generated for each agent. The evaluation process will be carried out using a performance score generated by the agent as a response to the human agent's correction of actions.

5. Description of Personal and Facilities

5.1. Work Breakdown Structure

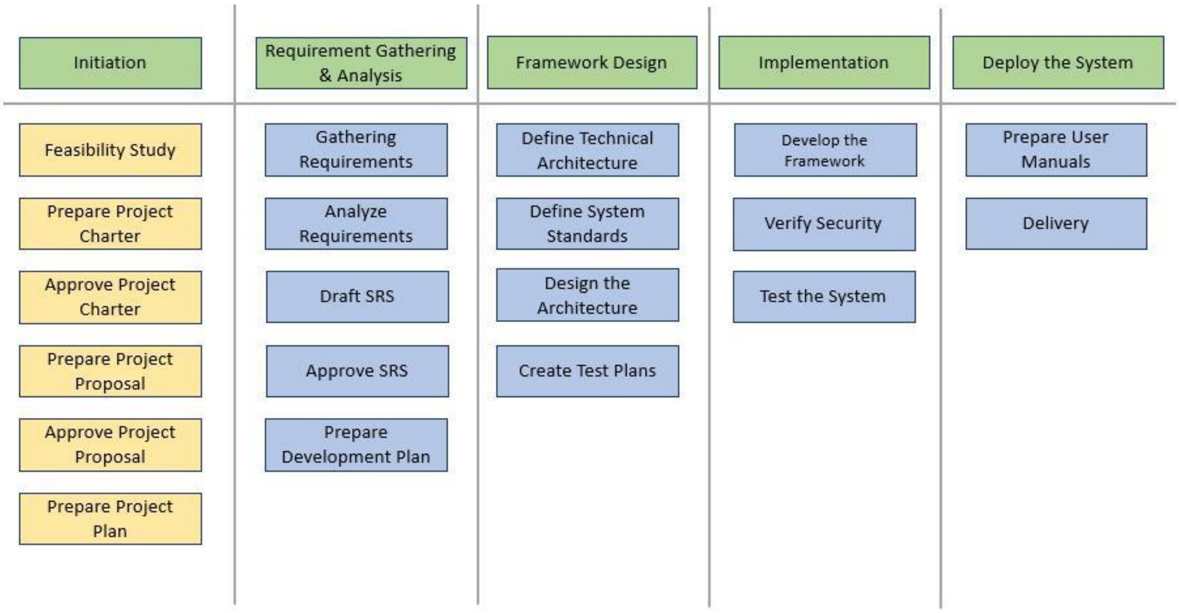

Figure 5.1.1: Work Breakdown Structure

5.2.Gantt Chart

Figure 5.2.1: Gantt Chart

5.3.Workload Allocation

Student ID	Name of the Student	Workload
IT16090804	Gunawardhana M.D.R.T.	Manages ongoing conversation dialogues with Machine Learning and improves with each conversation based on the feedback provided. Deep Reinforcement Learning will be used to create a self-learning conversation management system which will be pre-trained using existing conversation histories. A knowledge graph like Neo4j will be used to represent knowledge (structured data – entities, intent) and its relationship with each other.
IT18081794	Herath H.M.D.N.	Convert Natural Language into structured data by extracting intent, entities, and other structured information. Natural language processing modules will be implemented which support both Sinhala & English. A deep learning neural network will be implemented to extract important keywords like entities, intent, and actions from the given text input. The module will self-learn with each

		feedback and will be pre-trained.
IT17109536	Anjali R.P.D.N.	Convert voice sequences received into text format. The system will be implemented based on the Deep Speech 3 implementation by Baidu to support both English and Sinhala languages. Recurrent Neural Network will be trained using existing voice conversations using multiple GPUs. Separate language models will be implemented for each language.
IT16061880	Sampath G.A.D.M.	A state-of-the-art neural speech synthesis system. The system will be implemented based on the Deep Voice 3 and Wave NET implementation to support both English and Sinhala Languages. Recurrent Neural Network will be trained with a module containing an Encoder, Decoder & Convertor. The encoder converts textual features to an internal learned representation. Decoder decodes the learned representation while convertor creates the post-processing of audio to enable more human-like voice.

Table 5.2.1

6. Conclusion

This research proposal presented an automated screening tool to identify kids with autism capable of outperforming human agents in efficiency, performance, productivity, and availability with the highlight of Sinhala language support. Our approach is enabled by deep learning models with multi-GPU based large data set training combined with a data-driven conversational system. We believe this approach will continue to improve as we capitalize on increasing computing power and Sinhala language-based data sets in the future to revolutionize the screening process of kids with autism.

References

[1] J. D. Williams, K. Asadi, and G. Zweig, “Hybrid Code Networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning,” Feb. 2017.
[2] J. Li, W. Monroe, A. Ritter, M. Galley, J. Gao, and D. Jurafsky, “Deep Reinforcement Learning for Dialogue Generation,” Jun. 2016.
[3] V. Ilievski, C. Musat, A. Hossmann, and M. Baeriswyl, “Goal-Oriented Chatbot Dialog Management Bootstrapping with Transfer Learning,” Feb. 2018.
[4] N. Webb, “Rule-Based Dialogue Management Systems,” 2001.
[5] B. Peng et al., “Composite Task-Completion Dialogue Policy Learning via Hierarchical Deep Reinforcement Learning,” Apr. 2017.
[6] M. Gardner et al., “AllenNLP: A Deep Semantic Natural Language Processing Platform,” Mar. 2018.
[7] T. Bocklisch, J. Faulkner, N. Pawlowski, and A. Nichol, “Rasa: Open Source Language Understanding and Dialogue Management,” Dec. 2017.
[8] B. Hildebrandt, H. Rautenstrauch, and G. Sagerer, “Evaluation of spoken language understanding and dialogue systems,” Int. Conf. Spok. Lang. Process. ICSLP, Proc., vol. 2, pp. 685–688, 1996.
[9] A. Gorin, I. Langkilde-Geary, M. A. Walker, J. Wright, and H. W. Hastie, “Automatically Training a Problematic Dialogue Predictor for a Spoken Dialogue System,” J. Artif. Intell. Res., Jun. 2011.
[10] W. Ping et al., “Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning,” Oct. 2017.
[11] G. A. Rachel, S. L. Christina, V. S. Solomi, B. Ramani, P. Vijayalakshmi, and T. Nagarajan, “Development and analysis of various phone-sized unit-based speech synthesizers,” 2013 Int. Conf. Orient. COCOSDA Held Jointly with 2013 Conf. Asian Spok. Lang. Res. Eval. O-COCOSDA/CASLRE 2013, 2013.
[12] S. Lukose and S. S. Upadhya, “Text to speech synthesizer-formant synthesis,” 2017 Int. Conf. Nascent Technol. Eng. ICNTE 2017 - Proc., pp. 1–4, 2017.
[13] S. Rashmi, M. Hanumanthappa, and N. M. Jyothi, “Text-to-Speech translation using Support Vector Machine, an approach to find a potential path for human-computer speech synthesizer,” Proc. 2016 IEEE Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2016, pp. 1311–1315, 2016.

[14] M. Vasek, G. Rozinaj, and R. Rybarova, “Letter-To-Sound Conversation for Speech Synthesizer,” pp. 5–8, 2016.
[15] B. Ramani, A. J. M. P, P. Vijayalakshmi, and T. Nagarajan, “Voice Conversion-Based Multilingual to Polyglot Speech Synthesizer for Indian Languages,” 2013.
[16] S. Kannojia, “A Text to Speech synthesizer using acoustic unit-based concatenation for any Indian Language of Devnagari Script,” 2016.
[17] N. aziza Azis, R. M. Hikmah, T. V. Tjahja, and A. S. Nugroho, “Evaluation of Text-toSpeech Synthesizer for Indonesian Language Using Semantically,” pp. 978–979, 2011.
[18] R. Moumen, R. Chiheb, R. Faizi, and A. El Afia, “An NLP based text-to-speech. synthesizer for Moroccan Arabic,” 2017.
[19] A. W. . Black, H. Zen, and K. . Tokuda, “Statistical parametric speech synthesis,” Proc.\ ICASSP 2007, pp. 1229–1232, 2007.
[20] A. Hannun et al., “Deep Speech: Scaling up end-to-end speech recognition,” Dec. 2014.