ALIZA: SMART MIRROR AS AUTISTIC EDUCATION ASSISTANT

Pavidha Lojini Rajendran

(IT17131216)

B.Sc. (Hons) in Information Technology

Specializing in Software Engineering

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

September 2020

ALIZA: SMART MIRROR AS AUTISTIC EDUCATION ASSISTANT

Pavidha Lojini Rajendran

(IT17131216)

Dissertation submitted in partial fulfillment of the requirements for BSc (Hons) in Information Technology

Department of Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

September 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement 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 acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

……..… ……..

The above candidate has carried out research for the B.Sc. Dissertation under my supervision.

Signature of the supervisor: Date

……..… ……..

(Jesuthasan Alosius)

ABSTRACT

Autism is a neurodevelopmental syndrome that cause social communication and behavioral challenges that needs full-time care with therapies. According to WHO, one in 160 children has Autism Spectrum Disorders(ASD). This could be higher as the prevalence of ASD is far unknown in low and middle income countries. People with Autism has difficulty in understanding complex language, nonverbal communication as well as expressing their needs and emotions. However, early intervention can speed up child's development and reduce symptoms of autism at earliest. Speech therapy plays key part in early intervention programs.

Autism kids have strong visual processing skills so using pictures and technology can be more effective than speaking. Technological advances can potentially lead to novel and more effective treatment strategies and enhance quality of life for people with ASD and their families [1]. In this research, a smart mirror named "Aliza" is proposed with an objective of providing an interactive platform to grab their interest to boost their learning. Aliza provides verbal games with a speech recognition system in the backend to evaluate their performance which could help to create a significant level of improvement in their language skills.

Keywords: ASD, Verbal Trainer, Smart mirror, Speech recognition

ACKNOWLEDMENT

I wanted to express my sincere gratitude to every of the persons who gave me the opportunity to finish this project. An exemplary appreciation which I give to our project supervisor, Mr. Jesuthasan Alosius and Co-supervisor Mrs.Anjalie Gamage whose commitment to stimulating advice and guidance has helped for the project execution.

I would also like to thank Ms.Lakshika Udugama,Speech therapist at Lady Ridgeway for the suggestions and guidance that was provided to understand Autism and design the games. I would also like to express my deepest gratitude for the continuous love and encouragement of my caring parents and friends.

Finally, I would want to thank all SLIIT academic and non-academic staff, and every SLIIT classmate who supported me throughout the project.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ABBREVIATIONS

ASD	Autism spectrum disorder
AAC	Augmentative and Alternative Communication
PECS	Picture Exchange System
DTW	Dynamic Time Wrapping
HMM	Hidden Markov Model
MFCC	Mel-Frequency Cepstral Coefficient
CNN	Convolutional Neural Network
RNN	Recurrent Neural Nework
FFT	Fast Fourier Transform

1. INTRODUCTION

Autism spectrum disorder (ASD) is a complex developmental condition that involves persistent challenges in social interaction, speech and nonverbal communication, and restricted/repetitive behaviors [2]. Autism is heterogeneous condition that not all the people with autism has same profile. Even the communication disability varies from person to person. About 40% of children with ASD do not talk at all. About 25%–30% of children with ASD have some words at 12 to 18 months of age and then lose them. Others might speak, but not until later in childhood [3].

No cure found for autism but early interventions can reduce few symptoms and improve their abilities such as speech therapy for language and communication challenges. Speech pathologists perform evaluation before designing the therapy program for the kid. One of the speech therapies is Alterative Augmentative Communication (AAC) which is based on pictures and technologies. Aided AAC, in particular, would seem well suited to individuals with autism because it has been argued that they process visual information more easily than auditory information [4]. Usage of technologies like iPad, iPod and computers has already started changing the lives of people with ASD. High-tech system like speech generating device prompt phrases in computer generated voice when they tap on a picture or word. These methods are more effective than just speaking.

In this research, the smart mirror includes verbal trainer which provides simple speech activities. Furthermore, Children with ASD needs to be appreciated for what they do. “Praise is important to develop into a type of reinforcement because praise is a naturally occurring reinforcer” [5]. One of the characteristics of autism is having a fixation on one single activity and the inability to change course without extreme agitation, it is helpful that they know what to expect if they perform a task properly [6]. At the end of the completion of every tasks, Aliza provides positive reinforces to encourage them.

Aliza provides picture based activities such as to find objects in a picture, repeat sounds and words. As mentioned earlier, speech generating devices just speak phrases of word or pictures that is tapped by the kid. But the proposed solution takes the input voice to evaluate whether the kid attempts to repeat the word or make sound. According to their

response to the activity, progress level will be indicated. Through this system, it is expected to create a significant impact in language and communication skills of children with ASD.

1.1. Background Literature

Technology usage has prominently increased in treatment and research study of ASD. There are existing researches which used games to teach the autism kid and a speech recognition system within it. Following existing researches briefly describes about its features for autism kids and also included some of the literatures that researched about the speech command recognition system.

- **Computer Game to learn and enhance speech problems for Children with Autism [7]**

This system [7] has two applications that are management and learning application. Management application is for parents, teachers who are in charge of them to track their progress and their learning and the other application is to provide learning activities for the children, the child is encouraged to repeat the spoken word with the system. The voice signal will be analyzed to check how correctly they are pronouncing it. The voice recognition system in this app uses two different techniques for two languages. DTW and SPHINX4 has been used for Spanish and English respectively.

The preprocessing of the signal is done using MFCCs.23 ms size window is used for the fragmentation in the windowing algorithm and a 10ms time displacement. The correctly preprocessed MFCC compares with its equivalent one stored in the database to check which one is close to the obtained word in DTW algorithm. It calculates the time independent similarities between two vectors of the signal.

The other technique Sphinx4 is a framework based on statistical recognition using HMM and each associating HMM with single sound unit. In this framework, there are three modules that are Front end module which take the input signal, parameterizes and extract the features then the second module is linguist that translates any language with its pronunciation and structural information from one or more acoustic model into a SearchGraph. Finally, the third module Decoder generates the result from the information

from Linguist and Front-end. According to their research, even if the Sphinx gives issues, comparing the accuracy its performance is excellent.

- **Interactive Speech based Games for Autistic Children with Asperger Syndrome [8]**

Interactive Speech Based Game(ISBG) [8] includes two games implemented namely puzzle game and Picture Exchange Communication System(PECS) with Microsoft Speech technology. Speech Technology Programming Interface(SAPI) is developed by Microsoft that can decrease the code required to create an application for speech recognition and text-to-speech. SAPI consists Application Programming Interface(APII) and the Device Driver Interface(DDI).

The API is ISpEngine which has two components ISpRecoContext that is the main interface of speech recognition, receives a notification when speech recognition occurs and ISpRecoGrammar is to recognize the word that is defined. The DDI receives the audio from SAPI and return the recognition result. The system also has speech synthesis that handles the function using an API and DDI.

- **Noisy Speech Training in MFCC-based Speech Recognition with Suppression Toward Robot Assisted Autism Therapy [9]**

This research proposes a robot that has speech recognition system that can be used in noisy environment. In this study, noise suppression is applied to MFCC that averages the noise influence out. Then in each power energy the noise is suppressed in asymmetric noise suppression through asymmetric lowpass filter, half-wave rectifier and second-time asymmetric lowpass filter. After the process of noise suppression, the results are processed with cepstral mean normalization and dynamic range adjustment according to the process of MFCC.

Hidden Markov Model is used in this research for Speech recognition. This can capture the temporal variation and provide a framework to model relationship between acoustic feature and small set of phones. To calculate the output probability, Gaussian distribution is used and Baum-Welch algorithm is used to estimate the model parameter. In the test done to investigate the potential of this method, two experiments were tried out. In testing the performance of MFCC and MFCC with noise suppression using a clean training

set,MFCC with noise suppression provides the better accuracy of 92.61 than MFCC without noise supresssion.In the second experiment with noise training set, this method has provided a better accuracy than the previous experiment with clean training set and the Computational time is higher for MFCC with noise suppression compared to MFCC.

- **Improving ASR Systems for Children with Autism and Language Impairment Using Domain-Focused DNN Transfer Techniques [10]**

This study investigates about building and improving an automatic speech recognition system for children with Autism and language impairment. Limited data sources are available for children's speech set. To cope with the data limitations, this research has conducted two experiments. First to explore the strategies in transfer earning to optimize hyper parameters [10]. Second experiment is to explore the effect of augmenting the CSLU autism speech dataset with OSG kids' dataset.

In the first experiment, a time-delay neural network (TDNN) model was pre-trained on the LibriSpeech corpus dataset using the lattice-free maximum mutual information (MMI) ("chain") method and was built using the pre-configured scripts included in the Kaldi toolkit. The DNN architecture has a pre-trained and fixed linear discriminant analysis (LDA) input layer, six TDNN hidden layers, and a fully connected output layer with Batch normalization, L2 regularization and cross-entropy normalization employed. The research has used transfer learning technique to adapt LibriSpeech dataset. MFCC is used to extract the feature and The model is used to generate frame-level acoustic alignments for training LibriSpeech data. Weights for target model were copied from the LibriSpeech model and trained with the configurations over 5 epochs with an overall learning rate starting at 0.001 for epoch 0, decreasing linearly at each iteration to end at 0.0001. a weighted finite state transducer (WFST) language model that is built from the training data is used on decoding. LibriSpeech weights' performance was mentioned 80.75%.

In the second experiment, the effect of augmenting the CSLU Autism dataset is explored with the OGI Kids' dataset with training on the full K-3 set versus augmenting with each individual grade separately. Randomly selected 25% of each grade for a smaller K-3 set that matches the individual grades in quantity is used and the weights were transferred

from the LibriSpeech model, and retrained with a combination of the CSLU Autism data and each OGI Kids' segment.

1.2. Research Gap

Many researches are carried out to improve language and communication skills of children with ASD using technology. Technology based intervention is very effective in gaining attention and interest of the kid to focus on the tasks and skill development programs. Those researches have proven that it can produce positive outcomes. Augmentative Alternative Communication is a kind of assistive technology used by the speech pathologists. Mobile applications with AAC methods has already published. Most of these apps have enhanced the traditional way of Picture Exchange Communication System which is done using physical cards illustrating pictures. Along with PECS, computer generated voice read aloud the name of the card. It doesn't take the voice input of the kids to even check whether they have pronounced the word correctly whereas Aliza does.

Research A [11] uses series of card to send a simple message by combining them where the kid is not going to make an attempt to speak simple phrase or even repeat the word. Research B [12] is computer game to enhance the speech. This game has some of the functions that Aliza speech trainer provides. Compared to this game, Aliza doesn't have the feature to customize its activities. Research C [13] focused on building personal collection of words. Research D [14] is voice aid communication which is also fully customizable but lacking some other features like evaluating and logging the progress status. Below in Table 1, Features of Aliza is compared with other existing researches and mobile applications.

Table 1 : Comparison of existing researches

	Progress Report	Effective reinforces	Voice input	Choice-based activities	PECS	Customization
Research A	✗	✗	✗	✗	✓	✗
Research B	✓	✗	✓	✗	✓	✓
Research C	✗	✗	✓	✗	✗	✗
Research D	✗	✗	✓	✗	✓	✓
Otsimo	✗	✗	✓	✓	✗	✗
Proloquo2go	✗	✗	✗	✗	✓	✓
Aliza	✓	✓	✓	✓	✓	✗

1.3. Research Problem

Worldwide one in 160 children has an ASD. This estimate represents an average figure, and reported prevalence varies substantially across studies. Some well-controlled studies have, however, reported figures that are substantially higher [15]. ASD is a range of conditions with developmental disabilities such as social and interaction challenges, repetitive behavior. Disability to communicate or understand what people say can create anxiety and stress to the children with ASD. Rapid development in technology has enhanced the learning process of them. Technological tool for intervention can ensure efficient learning and speed up the skill development. AI and modern technology is already applied in treatment and diagnose of Autism and it has created an impact in developing the skills.

Even though there many other technologies like humanoid robots AI apps are already deployed to the market, still some therapist or schools have not integrated them to their teaching process. Humanoid robots are more expensive to be used in all the schools and mobile apps lack key functions to attract the attention and motivate the children. The smart mirror “Aliza” with activities is a novel solution to develop various skills and it is proven that children with ASD are more engaged with mirror and likes seeing their reflection.

1.4. Research Objectives

Main objective of this research is to improve the ability of language and communication of children with ASD.As previously mentioned, ASD varies person to person. Children with ASD who has limited speaking skills can benefit from the word repeating activities and recognizing objects and repeating sounds will at least make nonverbal kids try to speak.

- To build games which can help them to improve their language skills.
- To assess the performance of the kid using a speech recognition system.
- To make the kid use the learnt words in real world as well

1.4.1 Specific Objectives

- To identify the spoken word accurately regardless of noisy environment.
- To prompt reinforcers after completing task successfully.
- To ask for swapping another activity when kid seems less engaged.
- To design the game based on the characteristics of Autism.
- To generate a progress report based on how many words they succeeded to pronounce or how many objects they have identified during the game.

2. METHODOLOGY

2.1. Methodology

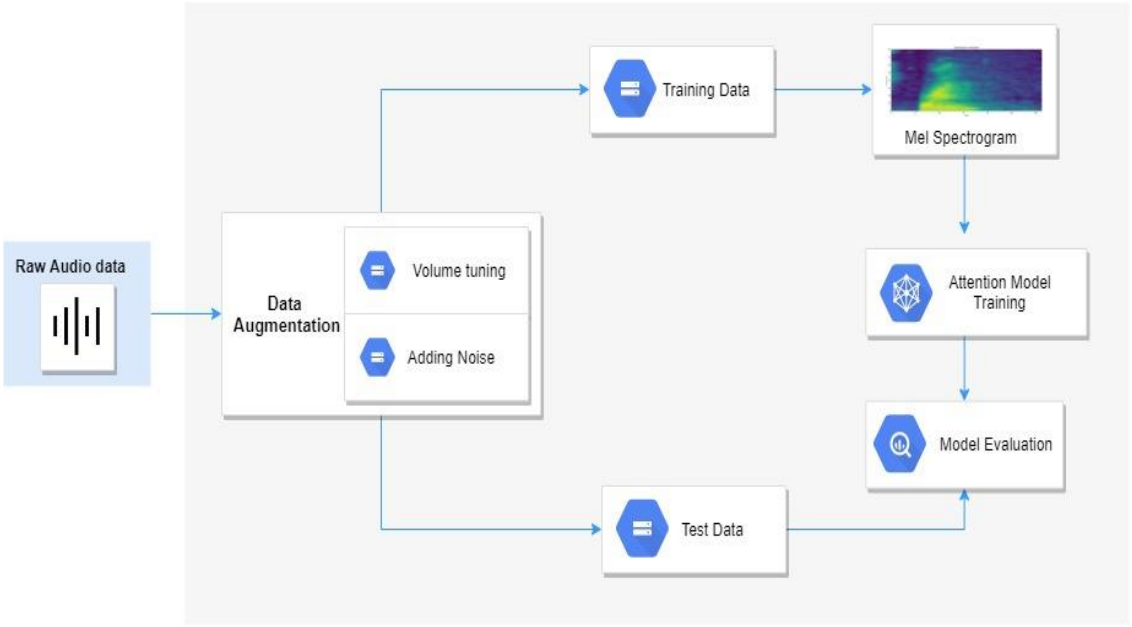

Figure 1:Model techniques

2.1.1 Dataset and Data visualization

Due to COVID-19 situation, the data collection process was postponed since the schools were closed. Tensor flow speech command dataset [16] was used for this research which contains 105,829 utterances, uttered by 2,618 speakers. These are a set of one second audio files recorded at 16kHz in .wav format. Each file contains single spoken English word. To help training the model to cope with noisy environment, Background noises were provided with the data set which contains longer audio clips which are recorded directly in the environments and some are generated mathematically. The details of the commands and noise are shown in Table 2.

Table 2: Commands in Dataset

Commands	Noise
Yes,No,Down,Up,Left,Right,On,Off,Stop,Go,Zero, One,Two,Three,Four,Five,Six,Seven,Eight,Nine,Backward,Forward,Learn,Visual,FollowBed,Bird,Cat,Dog,Happy,House,Marvin,Sheila,Tree,Wow	Doing the dishes Dude meowing Exercise bike Pink noise Running tap, White noise

Figure 2 illustrates the number of recordings of each spoken words. Speech and Language therapist at Lady Ridgeway has advised to use the words that will teach them the routinely used words. According to that, 20 words has been chosen from these command set to build the speech recognition system.

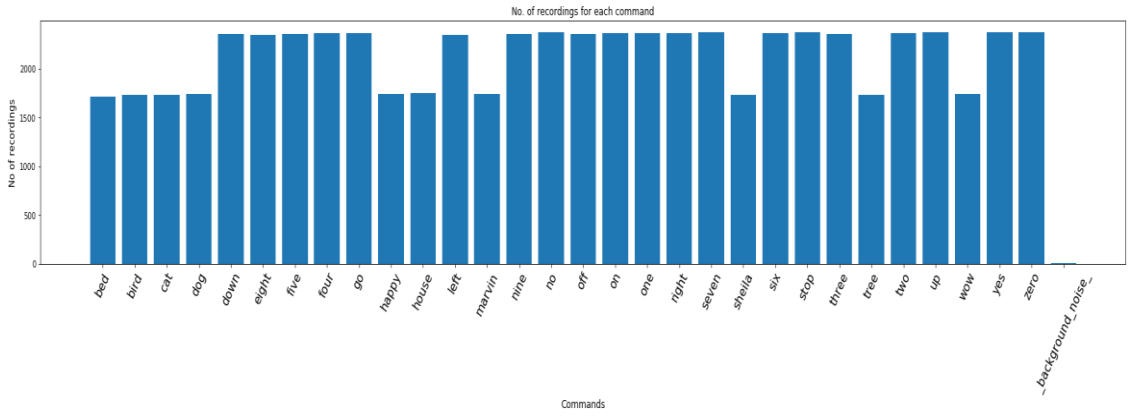

Figure 2:Number of Recordings

Following figure 3 visualize the amplitude of the signal over time for the word ‘Yes’ from the dataset.

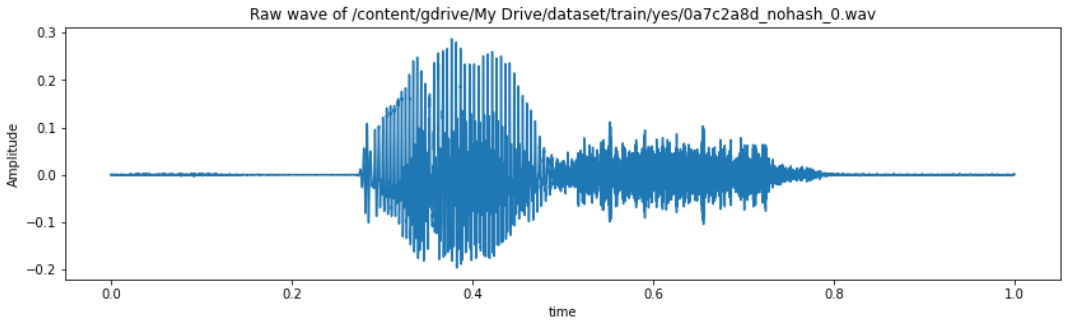

Figure 3:Raw signal of word 'Yes'

2.1.2 Pre Processing

Normally in nature audio signal is an analog signal which is a continuous representation over a time. It is need to convert the analog signal to digital signal to be processed efficiently and extract the useful information from the signal. A time domain analysis omits the rate of the signal depicted by frequency domain analysis whereas frequency domain analysis omits the sequence of the signal. As a solution for this, spectrogram denotes amplitude of a frequency at a time with use of colors. It is computed by applying Fourier Transform on overlapping windowed segments. Spectrogram is bunch of FFT stacked to represent various information. A mathematical operation is performed on frequencies to convert to mel scale which converts a spectrogram to mel scale spectrogram. Mel scale detects differences between lower frequencies that higher frequency.

In this model, the preprocessing is done within the keras layers to optimize the implementation. Kapre library facilitates to use mel spectrogram as a non-trainable layer.

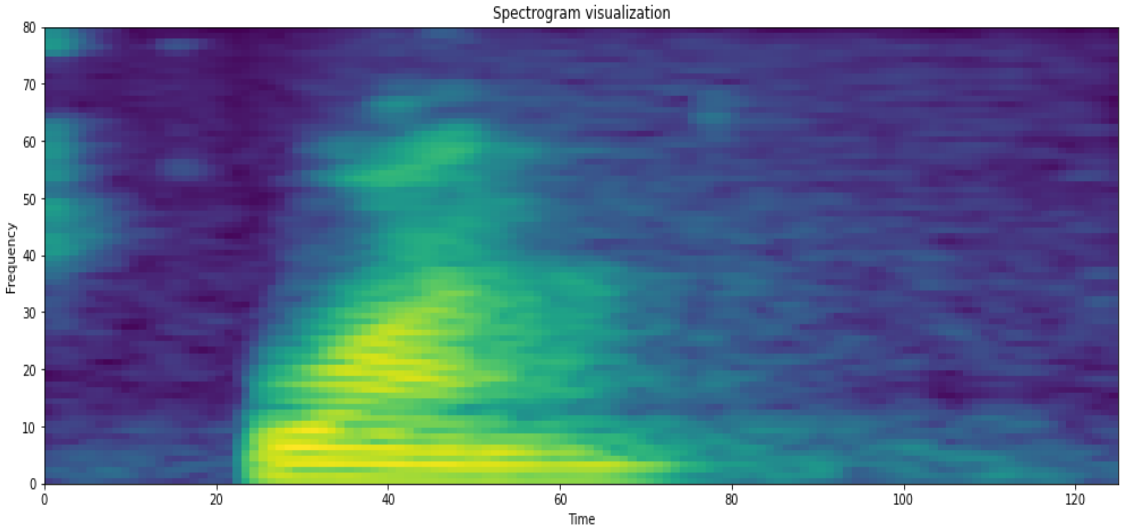

Figure 4:Mel Spectrogram

2.1.3 Data Augmentation

In the dataset, background noise was provided to increase the accuracy of the network. Volume augmentation is done by randomly increasing and reducing the volume of the audio. Figure 4 shows the signal representation of the clean audio file and Figure 5 depicts the representation of volume tuned audio where its amplitude reduced from 1 to 0.4.

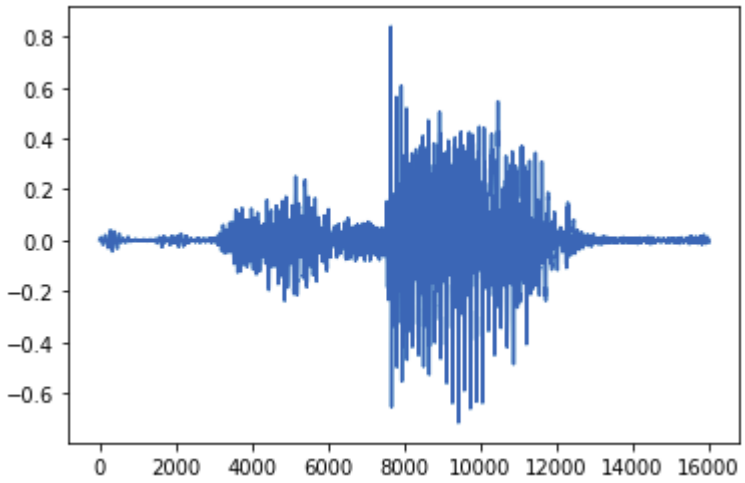

Figure 5:Signal representation of clean audio

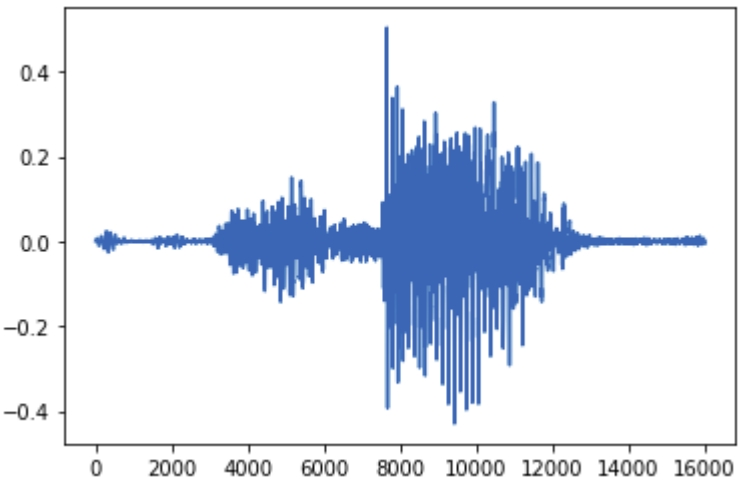

Figure 6:Signal representation of volume tuned audio

The clean data set was merged with the noisy data by adding noise to each audio file. Before the model training, among the five long audio clips of background noises, a 16khz of noise is extracted from randomly selected noise audio clip and merged it with the clean audio file.

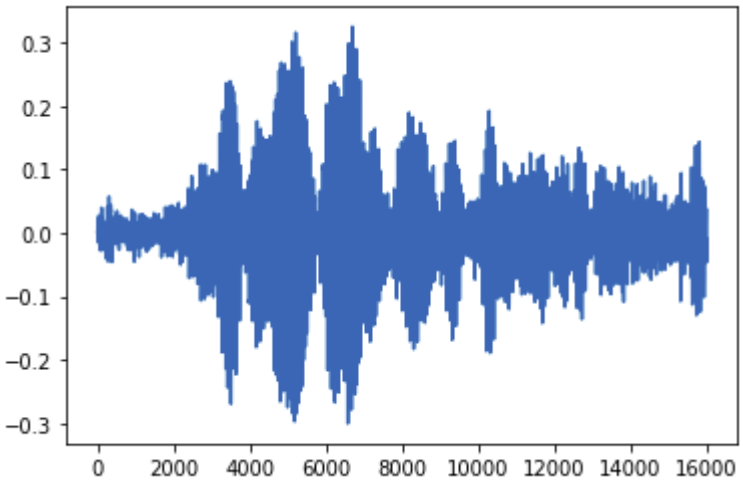

Figure 7:Signal representation of noise audio

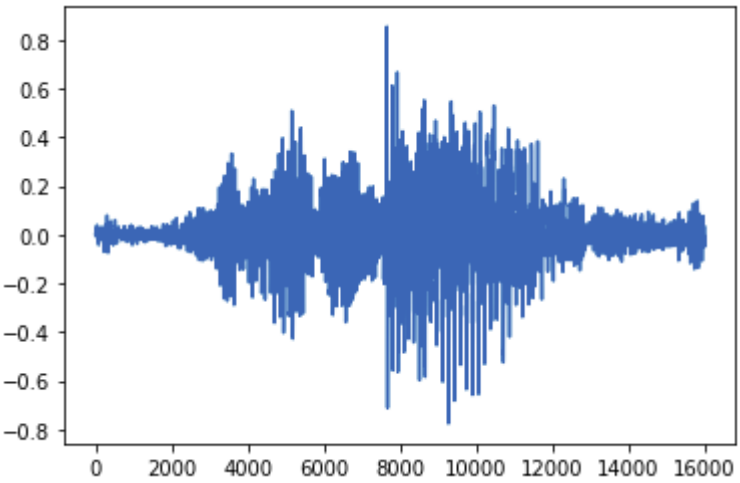

Figure 8:Signal representation of merged audio

2.1.4 Model training

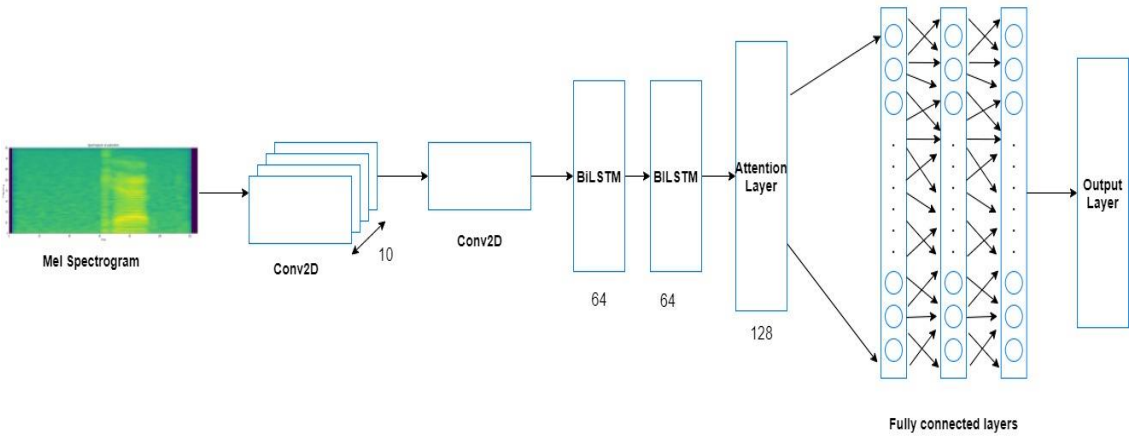

Figure 9 : Model architecture

For the microphone input, the model computes the Mel spectrogram in its non-trainable layer and two convolutional layers extract the local relations in the audio and two set of bidirectional LSTM that captures the two-way dependencies and the output is projected as a dense layer and utilized as query vector and finally the weighted output is fed into three fully connected layers. This is an attention model which provides a high accuracy to this dataset. A method [19] used in an existing methodology was enhanced in this literature and achieved more accuracy than the actual one by data augmentation technique.

Attention layer takes the middle vector of the LSTM output as bi directional LSTM contains the information of past and the future. Figure 10 depicts the attention weight of a random word in the list.

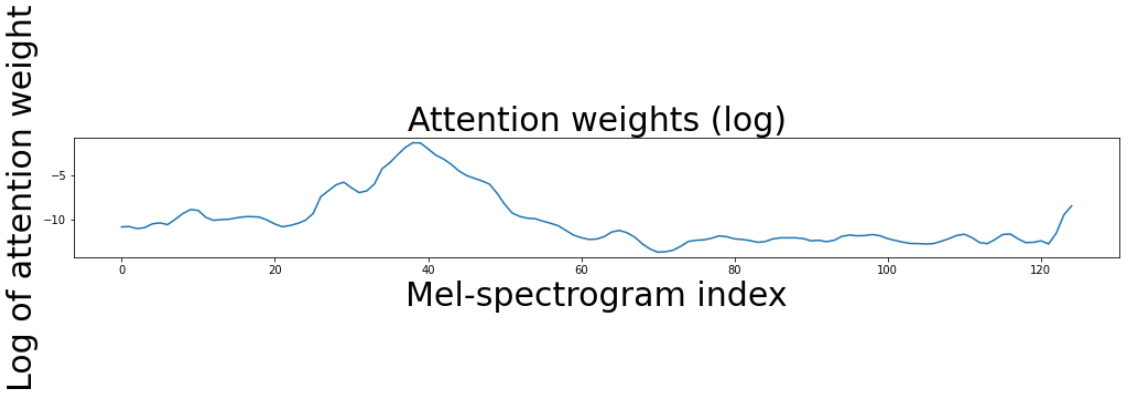

Figure 10: Attention weight plot

2.1.5 Game development

Game is the User interface where the children will be interacting while using the smart mirror. It is important to design the games in a more interactive way that can boost their interest on learning. Unity 2D is used to build the game component in Aliza smart mirror and it is built as a Linux application which resides inside magic mirror module. The game UI starts with a menu which has options to select different games such as pre-writing skills, numeric skills, focus game and speaking game along with login and report. Figure 11 show the UI of main menu.

Figure 11:Game main menu

Speaking game options takes the user to another menu where it shows three different games. First one is a game to learn Alphabet and then identifying object and finally arranging letters and repeating the word. The main menu of Speech games is shown below in Figure 12.

Figure 12:Main menu of verbal trainer

The first game prompts the alphabets and the user can move onto the next letter. The purpose of this game is to provide them the basics. Figure shows the UI of the game

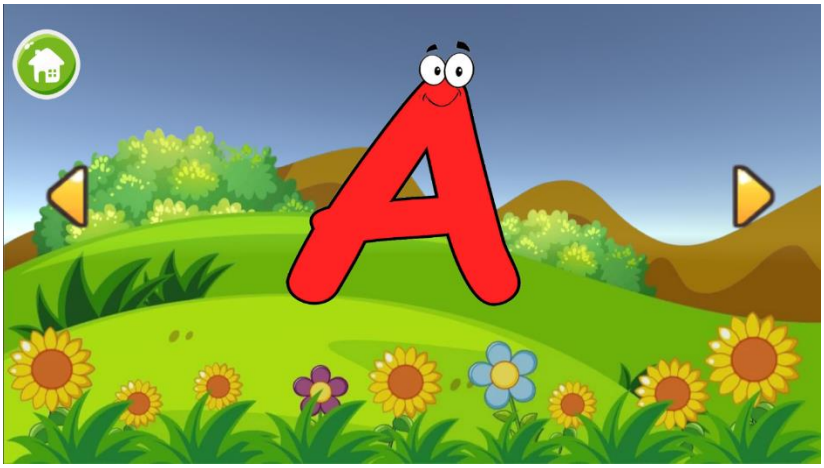

Figure 13:UI of alphabet game

The speech recognition system comes into play in the other two games that are object identification and spelling word game. Figure 14 visualizes the identification game UI. In this game, some of the objects move on the screen and child has to identify the object shown on the screen. There is a timer as well. The speech recognition system takes the voice input of the child and check how many words that the child has pronounced.

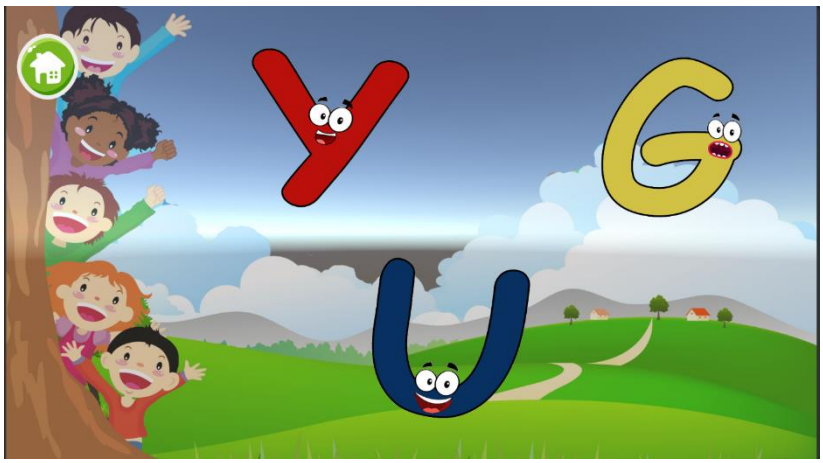

Figure 14:UI of object identification game

The last game is word spelling game. When the word block is dragged, a voice prompts the letter. If the user drag and drop a wrong letter in the wrong place, the block returns back to previous position. This continues until the word gets correctly arranged. A voice prompts the word after arranging it correctly and at the same time, the speech recognition system invokes and listen to the voice input of the kid. For the correct pronouncing of word, the game moves to next word with an attractive reinforce to encourage the child. Figure 15 show the UI of word spelling game for word 'Cat'. Only a few set of words are used in this game according to the advice of Speech therapist.

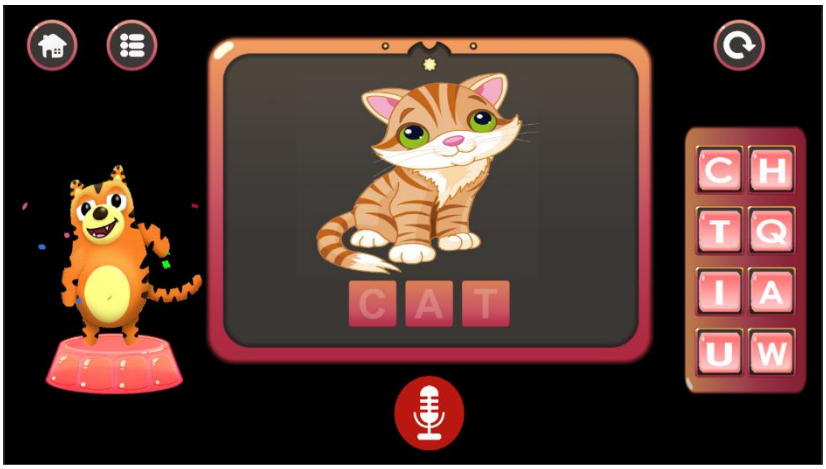

Figure 15:UI of word spell game

For each time the system generates a report for the person in charge for the kid to visualize the progress of the child. This report shows the number of words that a child successfully pronounced in between the given time and their progress level.

2.1.6 High-level architecture

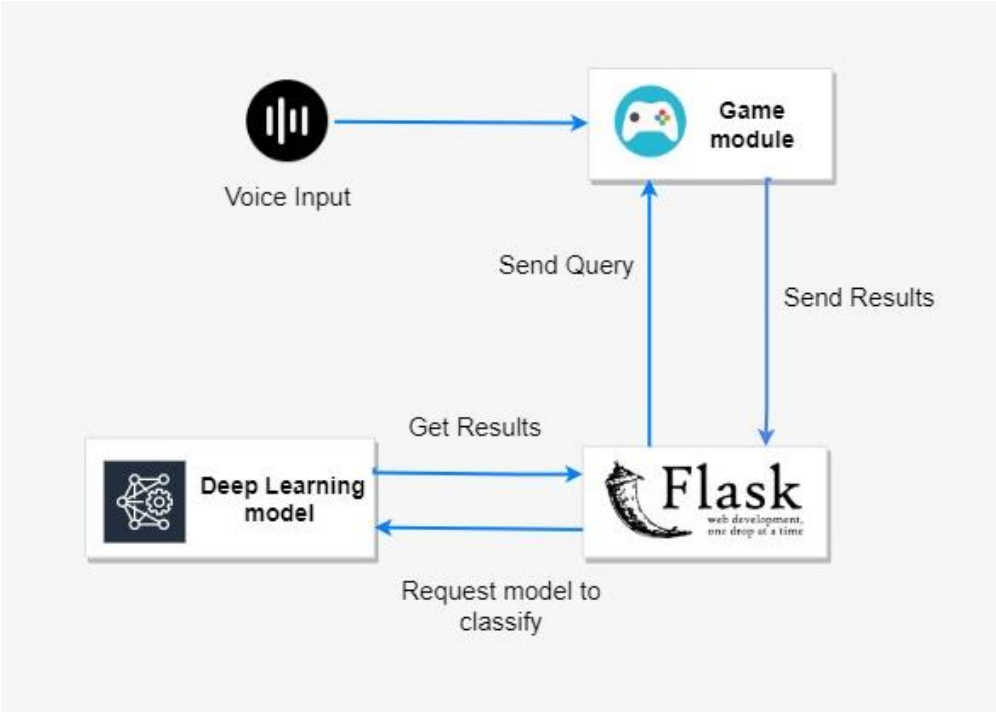

Figure 16: High level architecture of Verbal trainer

Figure 16 illustrates the overall architecture of the software module in Smart mirror. The Deep learning model is deployed to a Flask server and exposed as REST API. The API call is invoked from the Unity game and the speech data is sent as binary array to the Rest API for the processing to classify the word. Finally, the result is sent to the Database and the game screen. The processed Result in the database is used to generate the report for the progress of the child.

2.1.7 Smart mirror Hardware development

Two-way mirror with 70% reflection on top of LED monitor, covered with a wooden frame to hold them together. The LED monitor is to display the program. These two aesthetic components are 20 inches. A speaker and web camera in-built with microphone is connected to the Raspberry Pi for voice recognition and emotion recognition. Figure 17 illustrates the architecture of the system. Raspberry Pi is the brain of smart mirror which has the processing capability with low processing power. It is booted up with Linux 64-bit OS which is configured to run Magic mirror module during the startup. Inside the Magic mirror module, our application resides as a game that could be played by the user.

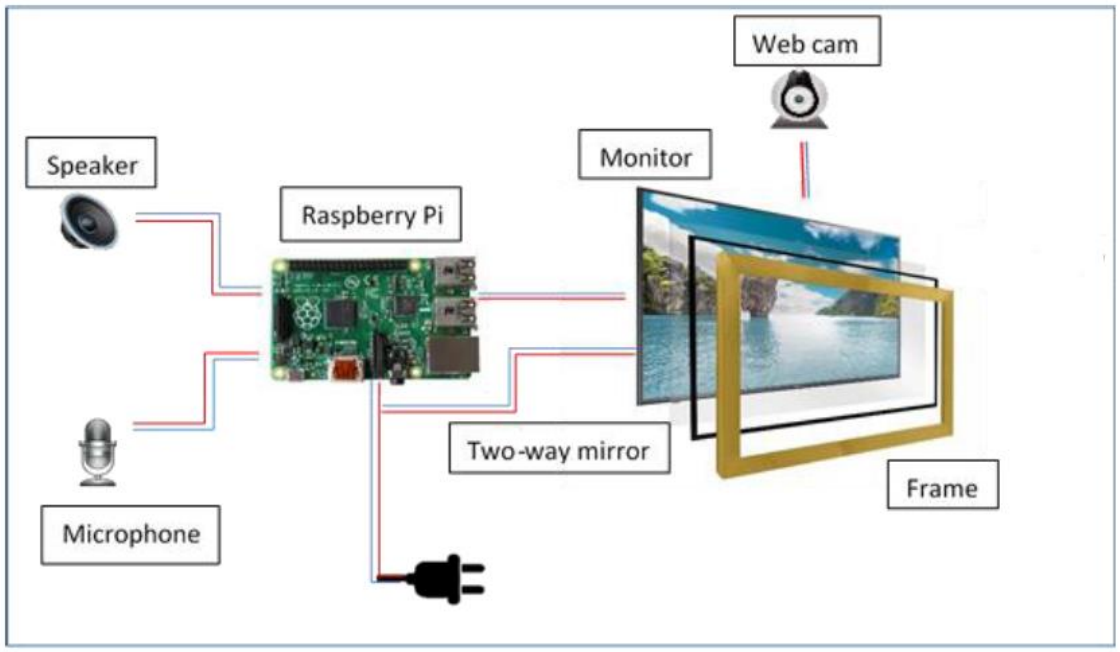

Figure 17:Hardware architecture diagram

In the process of booting Raspberry Pi, first the sd card was formatted using win32 disk imager and the Raspbian OS 32-bit was booted using the Raspberry Pi Imager that helps to easily install OS. A file inside the boot folder was created to setup the Wi-Fi configurations. To use the raspberry pi through the laptop, display, connected Raspberry pi using SSH with its IP address. VNC server is installed to remotely connect to the raspberry pi.

2.2. Commercialization Aspects of the Product

As indicated by a report by Fortune Business Insights, named, "Autism Spectrum Disorder Therapeutics Market: Global Market Analysis, Insights and Forecast, 2018-2026," the worldwide market is probably going to reach US$ 4,612.1 Mn by 2026 inferable from the rising rate of autism spectrum disorder around the world. According to the report, ASD therapeutics market was worth US$ 3,293.0 Mn in 2018. Considering the expanding awareness programs with respect to this specific issue and less accessibility of treatment alternatives, the worldwide market is relied upon to ascend at a moderate CAGR of 4.3% during the time frame (2019-2026) [17]. Figure 18 depicts an overview of the above mentioned report. Providing a solution like Aliza with the modern trending technology can be significantly commercialized as it shows an increase market need.

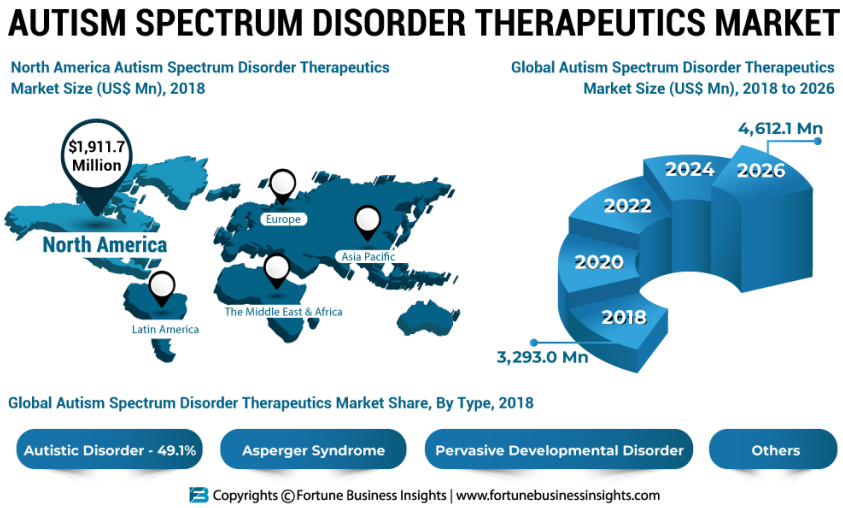

Figure 18:Autism Therapeutic Market

Children with ASD are often self-absorbed and seem to exist in a private world in which they have limited ability to successfully communicate and interact with others [18]. Intervention programs to improve their communication skills is vital need that can help them to socially interact with people and reduce social anxiety. Aliza helps to improve one of the major challenges that ASD children has. The sub component verbal trainer provides interactive game which can keep them more interested and engaged.

The unique aspect that can promote this product is having an AI evaluation system to track their progress and generate report. Verbal trainer resides inside the magic mirror

module has a speech command recognition system built on deep learning. This system can take the voice input of Autism child and check whether they have correctly pronounced the word or not in between a given time. Other application exists in the market doesn't contain an evaluation system that Aliza provides.

2.3. Testing and Implementation

2.3.1 Implementation

Kapre library is used to preprocess the audio signal to mel spectrogram with 80 band mel scale and,1024 discrete Fourier transform and 128 points of hop size. Then it is trained in 7 more layers. Figure 19 shows the model summary.

Layer (type)	Output Shape	Param #	Connected to
input (InputLayer)	[(None, None)]	0	
reshape (Reshape)	(None, 1, None)	0	input[0][0]
mel_stft (Melspectrogram)	(None, 80, None, 1)	1091664	reshape[0][0]
mel_stft_norm (Normalization2D)	(None, 80, None, 1)	0	mel_stft[0][0]
permute (Permute)	(None, None, 80, 1)	0	mel_stft_norm[0][0]
conv2d (Conv2D)	(None, None, 80, 10)	60	permute[0][0]
batch_normalization (BatchNorma	(None, None, 80, 10)	40	conv2d[0][0]
conv2d_1 (Conv2D)	(None, None, 80, 1)	51	batch_normalization[0][0]
batch_normalization_1 (BatchNor	(None, None, 80, 1)	4	conv2d_1[0][0]
squeeze_last_dim (Lambda)	(None, None, 80)	0	batch_normalization_1[0][0]
bidirectional (Bidirectional)	(None, None, 128)	74240	squeeze_last_dim[0][0]
bidirectional_1 (Bidirectional)	(None, None, 128)	98816	bidirectional[0][0]
lambda (Lambda)	(None, 128)	0	bidirectional_1[0][0]
dense (Dense)	(None, 128)	16512	lambda[0][0]
dot (Dot)	(None, None)	0	dense[0][0] bidirectional_1[0][0]
attSoftmax (Softmax)	(None, None)	0	dot[0][0]
dot_1 (Dot)	(None, 128)	0	attSoftmax[0][0] bidirectional_1[0][0]
dense_1 (Dense)	(None, 64)	8256	dot_1[0][0]
dense_2 (Dense)	(None, 32)	2080	dense_1[0][0]
output (Dense)	(None, 36)	1188	dense_2[0][0]

Figure 19:Model summary

The model was trained for 60 epochs and the best model is saved with best accuracy of validation set. If the validation accuracy continues to be same for 10 epochs the training stops. The optimizer ‘adam’ is used to change the weights and bias terms in the network. Figure 20 shows the early stopping of model while training.

```
Epoch 00044: val_sparse_categorical_accuracy did not improve from 0.95267
2651/2651 - 83s - loss: 0.1382 - sparse_categorical_accuracy: 0.9610 - val_loss: 0.1973 - val_sparse_categorical_accuracy: 0.9507
Changing learning rate to 6.400000000000001e-05
Epoch 45/60

Epoch 00045: val_sparse_categorical_accuracy did not improve from 0.95267
2651/2651 - 83s - loss: 0.1346 - sparse_categorical_accuracy: 0.9623 - val_loss: 0.1986 - val_sparse_categorical_accuracy: 0.9498
Changing learning rate to 6.400000000000001e-05
Epoch 46/60

Epoch 00046: val_sparse_categorical_accuracy did not improve from 0.95267
2651/2651 - 83s - loss: 0.1372 - sparse_categorical_accuracy: 0.9620 - val_loss: 0.1906 - val_sparse_categorical_accuracy: 0.9514
Changing learning rate to 6.400000000000001e-05
Epoch 47/60
Restoring model weights from the end of the best epoch.

Epoch 00047: val_sparse_categorical_accuracy did not improve from 0.95267
2651/2651 - 83s - loss: 0.1349 - sparse_categorical_accuracy: 0.9622 - val_loss: 0.1946 - val_sparse_categorical_accuracy: 0.9496
Epoch 00047: early stopping
```

Figure 20: Snip of early stopping of the model

Following are the files used in the implementation of this model

- SpeechCmdRecognitionWithAug.ipynb
- SpeechGeneratorWithAug.py
- FlaskApp.py

Here the SpeechGenerator.py consists the code of adding noise to the data set and SpeechCmdRecognition notebook does the model training.

Finally, the model is converted to tensorflowlite for optimization to deploy to Raspberry pi. Then a flask app is built in FlaskApp.py to continuously stream the microphone input and predict the word.

2.3.2 Testing

A product's success is depending on its quality and reliability. It is important to test the product before releasing it. After completing the implementation, testing was carried out to check for bugs in the system. There are four components in this system. They are writing mentor, Math tutor, Verbal Trainer and attentiveness tracker. Integration testing was done after integrating the four components. Below in this section some of the test cases that were used in the testing phase I discussed. Found bugs were corrected and again integrated with the system.

- Unit Testing

Unit testing is done by appointing the planning cycle into sub parts known as units. During this stage, every component of the four project group individuals was tried independently. The components were later coordinated as one unit at long last to jump toward mix and framework testing. Every component of the Aliza has been created and tried for its particular usefulness and this testing technique affirms whether the component have accomplished their purpose characterized at an opportune time during the requirement analysis stage.

For the initial phase, the Speech recognition was experimented with the Tensorflow speech command dataset which is a dataset of audio recordings from the adults for selected words. This data set contains 105,829 utterances, uttered by 2,618 speakers These are a set of one second audio files recorded at 16kHz in .wav format. In second phase, the components were tested with the in-built microphone in the laptop in order to further enhance the performance and accuracy. During both the component testing phases, Speech recognition component was unit tested using White-Box testing approach against a set of defined test cases.

Table 3:Recognition of Word 'Cat' Test case

Test Case ID	**1**
Test Case Name	Recognize word ‘Cat’ in isolated area
Test Input Data	Pronounce a word from the selected list of words
Expected Output	Text form of the spoken word ‘Cat’
Actual Output	Printed the word ‘Cat’

Table 4:Recognition of Word 'Dog' Test case

Test Case ID	**2**
Test Case Name	Recognize word ‘Dog’ in noisy area
Test Input Data	Pronounce the word ‘Dog’ from a noisy area
Expected Output	Text form of the spoken word ‘Dog’
Actual Output	Printed the word ‘Dog,Up’

Table 5:Recognition of Word 'Cup' Test case

Test Case ID	**3**
Test Case Name	Identify the unknown word
Test Input Data	Pronounce ‘Cup’ that is not in the selected words list
Expected Output	Label the word as ‘Unknown’
Actual Output	Printed the word ‘Unknown’

- Integration Testing

In this testing phase we have combined the units tested individual modules together and tested as an integrated module. The objective of integration testing was to make sure that the integrated system runs smoothly, even if the components or the units of the software works fine individually. For integration testing phase the Black-Box testing approach was followed.

3. RESULTS & DISCUSSION

3.1. Results

After the data augmentation to the model, the model achieved a test accuracy of 91% for noisy data and 94% for clean dataset. Analyzing the training and validation accuracy it is visible that after performing the data augmentation overfitting in the model was also tackled.

```
Evaluation scores:
Metrics: ['loss', 'sparse_categorical_accuracy']
Train: [0.15777570009231567, 0.9563136696815491]
Validation: [0.17972245812416077, 0.9533762335777283]
Test: [0.19719092547893524, 0.944479763507843]
TestWithNoise [0.291598379611969, 0.9183098673820496]
```

Figure 21: Evaluation scores of model with data augmentation

Figure 22 shows the evaluation score of the trained model with clean data set and the model achieved a test accuracy rate of 80% with noisy dataset.

```
Evaluation scores:
Metrics: ['loss', 'sparse_categorical_accuracy']
Train: [0.04987237602472305, 0.9886953234672546]
Validation: [0.24248231947422028, 0.9442323446273804]
Test: [0.23336762189865112, 0.9412994384765625]
TestWithNoise [0.8443034887313843, 0.8069968223571777]
```

Figure 22:Evaluation score of model with clean data

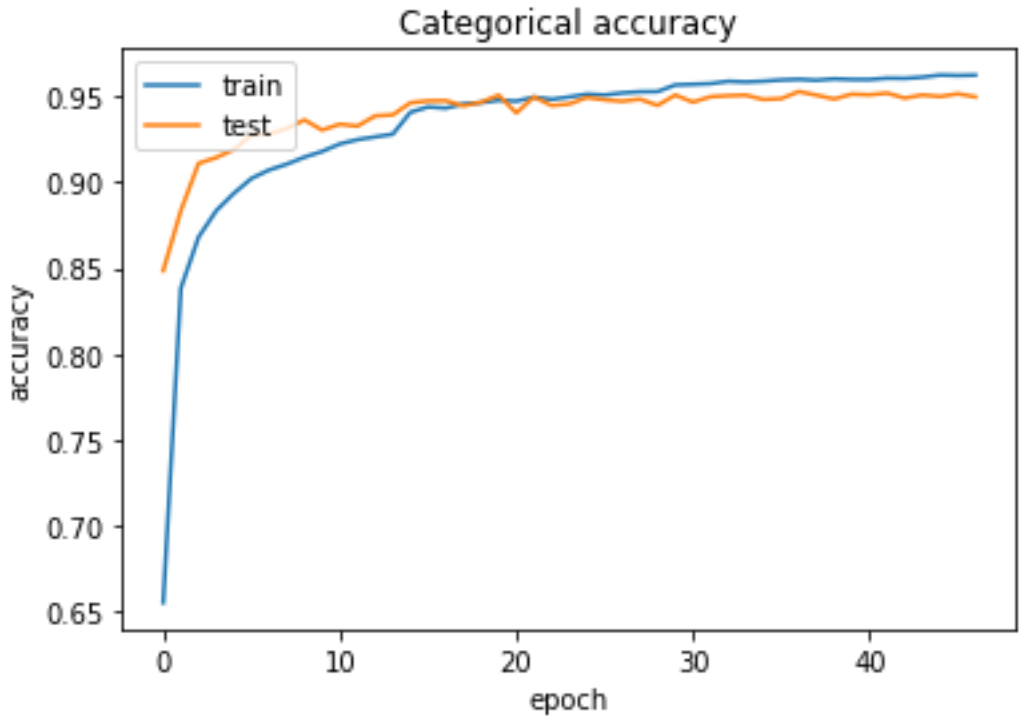

Figure 23:Accuracy plot

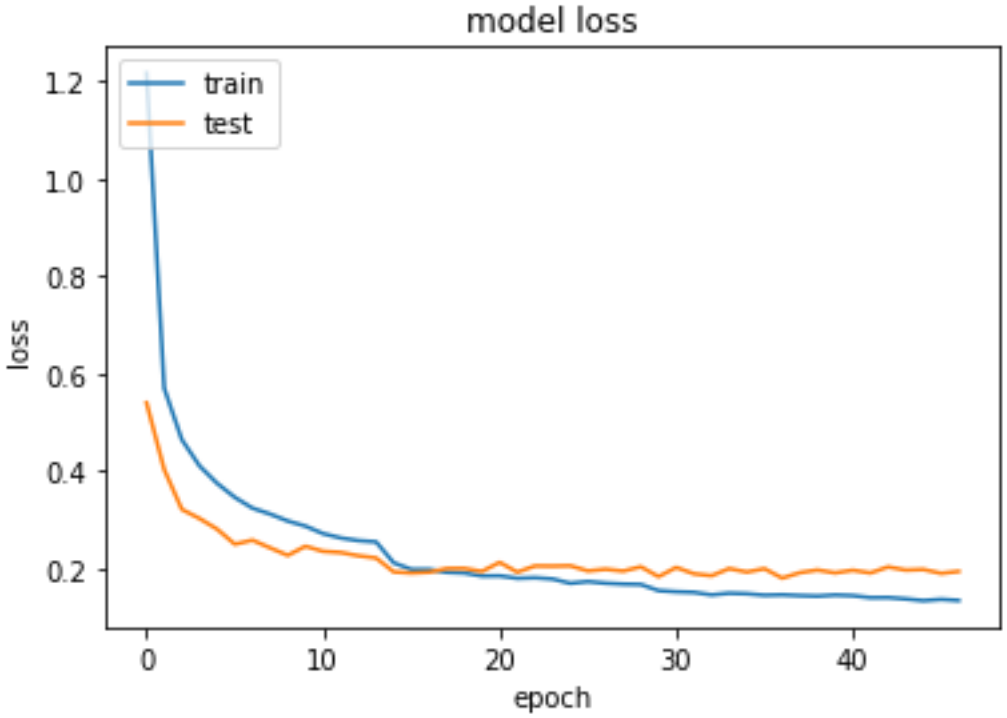

Figure 24:Model loss plot

The confusion matrix shows the predicted labels of the test set below in figure 22.

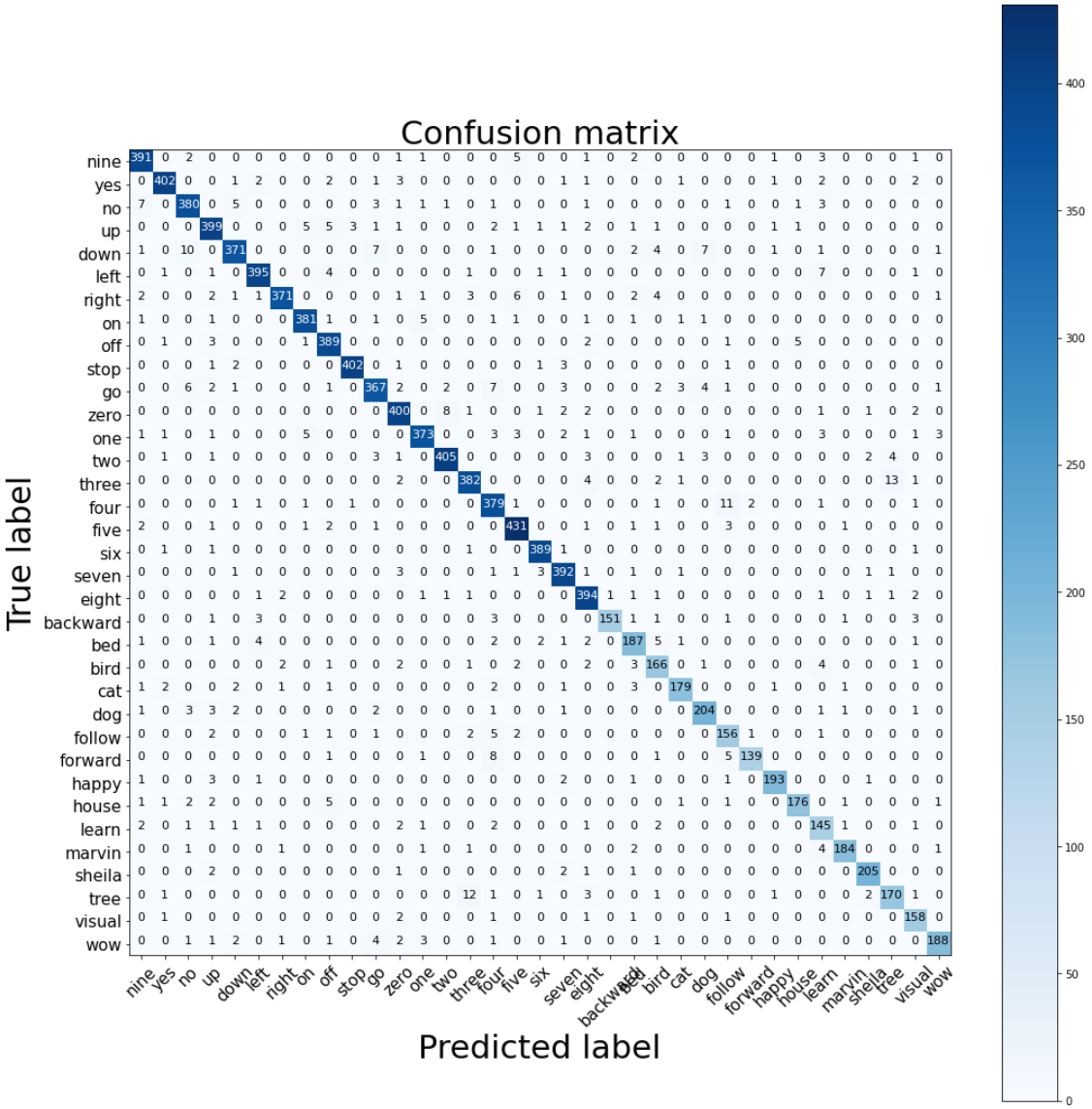

Figure 25:Confusion matrix of model

The pre-trained CNN model has achieved the test accuracy of 90%. Two experiments were conducted to the test the results of the speech component.

- Testing in an isolated environment

Table 6: Experiment result in isolated area

Phrases	Success Rate (%)	No Of Test Run Audio
House	60	5
Cat	100	5
Dog	100	5
Happy	80	5
Up	100	5

For this experiment, five words were pronounced by 2 normal kids and 3 autism kids to compute the success rate of the speech recognition system. In this test, the system has identified the most of the words pronounce by them but the small two or 3 letter words were easily recognized and a fault rate was detected in the 5 letter long words.

- Testing in noisy environment

Table 7: Experiment result in noisy environment

Phrases	Success Rate (%)	No Of Test Run Audio
House	60	5
Cat	100	5
Dog	100	5
Happy	80	5
Up	100	5

For this experiment, the same 2 normal kids and the 3 autism kids were used to pronounce the word. In this test, the system has recognized only a few words. The noisy environment had the noise of vehicles as well as the sounds from other kids in the classroom. therefore, the system was trying to listen to the sound from the other kids and resulted the wrong words.

3.2. Research Findings and Discussion

There are existing researches which has tried to implement games with speech recognition system. Those systems have used existing technologies such as SPHINX4

and Microsoft speech technology which is built for adult speech and not specific to the autism kids which is a drawback to the system. According to the literature review, they have used some advanced technique to obtain the accuracy and overcome the data scarcity for autism speech. All these techniques are implemented for long speeches which can cause to high computational time and less accuracy than the command classifying approach.

In this system, the backend of the game is approached as command recognition which will classify the correct label of word which can add more accuracy in identifying the pronounced word. During the implementation, in the first phase, the audio signal was preprocessed and fed to the model for training. This model had four convolutional layers and used adam optimizer for the best results.even though this model was able to achieve 90% of the accuracy during model training,it predicted the word wrong. Therefore, in the second phase after the preprocessing, the feature extraction step was added to the techniques. For the feature extraction, MFCC is used which is more human representative. Considering MFCC image data, approached the solution as image classification and used a Convolutional neural network with eight layers. This model has achieved a test accuracy of 70%. After further research on the context, an attention model using RNN [19] has achieved a significant accuracy of 94% which has given a better result while testing it with microphone input after deploying to flask server. This model starts by computation of mel spectrogram as non-trainable layer. Kapre library provides the facility to use mel spectrogram as layer.

In addition, according to an existing literature [20], Data augmentation technique was utilized to improve noise robustness. The dataset has already provided long audio clips of noises recorded from the environment and mathematically generated. These noise files were merged with the audio files to train the network to recognize the word even in a noisy environment. As a result of the technique, model achieved test accuracy of 91% for the noisy data.

4. CONCLUSION

With the increase in autism children [21] there is a lack of resource to cater to the needs of all. Nevertheless, this research paper states how ‘Aliza’ smart mirror sets a new standard among the training and teaching methods. Therefore, this research holds training and teaching methods based on game activities which are helps to increase interest and focus of autism students and it accommodates basic education for age groups of five to ten aged children. The sub component Verbal trainer helps to improve the speech and language abilities of ASD children. The use of speech recognition system with the game adds more value to the component. Evaluating them throughout the game makes the care taker observe their progress. Considering the features of other existing applications, this proposed system could create a significant improvement in their learning.

As the future enhancement for this system, it is expected to provide support for local languages such as Tamil and Sinhala. In the teaching process of Autism children, it is important to keep them engaged in an interactive way and also make them interactive with the environment as well rather than being stick to the technology. Therefore, the games should be designed more fun to play and learn.

REFERENCES

[1] Bölte, S., Golan, O., Goodwin, M. and Zwaigenbaum, L. (2010). What can innovative technologies do for Autism Spectrum Disorders?. *Autism*, 14(3), pp.155-159.

[2] Psychiatry.org. (2020). *What Is Autism Spectrum Disorder?*. [online] Available at: https://www.psychiatry.org/patients-families/autism/what-is-autism-spectrum-disorder [Accessed 20 Feb. 2020].

[3] Centers for Disease Control and Prevention. (2020). *Signs & Symptoms | Autism Spectrum Disorder (ASD) | NCBDDD | CDC*. [online] Available at: https://www.cdc.gov/ncbddd/autism/signs.html [Accessed 22 Feb. 2020].

[4] Iacono, T., Trembath, D. and Erickson, S. (2016). The role of augmentative and alternative communication for children with autism: current status and future trends. *Neuropsychiatric Disease and Treatment*, Volume 12, pp.2349-2361.

[5] Long, S. (2020). *Teaching Tip: Positive Reinforcement! - The Autism Helper*. [online] The Autism Helper. Available at: http://theautismhelper.com/teaching-tip-positive-reinforcement/ [Accessed 22 Feb. 2020].

[6] About Autism. (2020). *Positive Reinforcement - About Autism*. [online] Available at: http://aboutautism.net/positive-reinforcement/ [Accessed 22 Feb. 2020].

[7] M. Frutos, I. Bustos, B. Zapirain and A. Zorrilla, "Computer game to learn and enhance speech problems for children with autism", *2011 16th International Conference on Computer Games (CGAMES)*, 2011. Available: 10.1109/cgames.2011.6000340 [Accessed 19 September 2020].

[8] A. Alqahtani, N. Jaafar and N. Alfadda, "Interactive speech based games for autistic children with Asperger Syndrome", *The 2011 International Conference and Workshop on Current Trends in Information Technology (CTIT 11)*, 2011. Available: 10.1109/ctit.2011.6107947 [Accessed 26 September 2020].

[9] S. Attawibulkul, B. Kaewkamnerdpong and Y. Miyanaga, "Noisy speech training in MFCC-based speech recognition with noise suppression toward robot assisted autism therapy", *2017 10th Biomedical Engineering International Conference (BMEiCON)*, 2017. Available: 10.1109/bmeicon.2017.8229135 [Accessed 26 September 2020].

[10] R. Gale, L. Chen, J. Dolata, J. Santen and M. Asgari, "Improving ASR Systems for Children with Autism and Language Impairment Using Domain-Focused DNN Transfer Techniques", *Interspeech 2019*, 2019. Available: 10.21437/interspeech.2019-3161 [Accessed 26 September 2020].

[11] Raja, Pravind & Saringat, Mohd & Mustapha, Aida & Zainal, Abidah. (2017). Prospect: A Picture Exchange Communication System (PECS)-based Instant Messaging Application for Autism Spectrum Condition. IOP Conference Series: Materials Science and Engineering. 226.

[12] M. Frutos, I. Bustos, B. G. Zapirain and A. M. Zorrilla, "Computer game to learn and enhance speech problems for children with autism," 2011 16th International Conference on Computer Games (CGAMES), Louisville, KY, 2011, pp. 209-216.

[13] Wilson, Cara & Brereton, Margot & Ploderer, Bernd & Sitbon, Laurianne. (2018). MyWord: enhancing engagement, interaction and self-expression with minimally-verbal children on the autism spectrum through a personal audio-visual dictionary. 106-118.

[14] I. Torii, K. Ohtani, N. Shirahama, T. Niwa and N. Ishii, "Voice output communication aid application for personal digital assistant for autistic children," 2012 IEEE/ACIS 11th International Conference on Computer and Information Science, Shanghai, 2012, pp. 329-333.

[15] Who.int. (2020). *Autism spectrum disorders*. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders [Accessed 25 Feb. 2020].

[16]"speech commands | Tensor Flow Datasets", *TensorFlow*, 2020. [Online]. Available: https://www.tensorflow.org/datasets/catalog/speech_commands. [Accessed: 26- Sep- 2020].

[17] "Autism Spectrum Disorder Therapeutics Market Size, Growth, Analysis, Insights and Forecast 2019-2026 | Medgadget", *Medgadget.com*, 2020. [Online]. Available: https://www.medgadget.com/2019/12/autism-spectrum-disorder-therapeutics-market-size-growth-analysis-insights-and-forecast-2019-2026.html. [Accessed: 29- Sep- 2020].

[18] "Autism Spectrum Disorder: Communication Problems in Children", *NIDCD*, 2020. [Online]. Available: https://www.nidcd.nih.gov/health/autism-spectrum-disorder-communication-problems-children. [Accessed: 29- Sep- 2020].

[19] D. C. de Andrade, S. Leo, M. L. Da S. Viana and C. Bernkopf, "A neural attention model for speech command recognition", 2018, [online] Available: https://arxiv.org/abs/1808.08929

[20] A. Pervaiz et al., "Incorporating Noise Robustness in Speech Command Recognition by Noise Augmentation of Training Data", *Sensors*, vol. 20, no. 8, p. 2326, 2020. Available: 10.3390/s20082326.

[21] The Autism Community in Action, "Autism Statistics & Cost", [Online]. Available: https://tacanow.org/autism-statistics/ [Accessed: 22 Feb 2020].