International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET122936

Virtual talk assistance for Dumb and Deaf

Suchethana H C, Arun Kumar P
Assistant Professor, ISE Department, JNNCE, Shivamogga, Karnataka, India

Article Info
Volume 9, Issue 3
Page Number : 96-105
Publication Issue :
May-June-2022
Article History
Accepted : 01 May 2022
Published: 13 May 2022

ABSTRACT
For the Deaf and Dumb community, the use of Information and Communication Technology has increased the ease of life for them. Deaf and dumb people communicate with help of sign language to pass their messages to each other. In this way, they can't express their ideas the exact way, they want. The implementation of Speech To Text and Text To Speech technique for deaf and dumb people to make them communicate better. Here the compared elaborated research work done in the same field which may be helpful in identifying the drawbacks as well as methods to improvise the present technology. It presents a detailed methodology used in numerous research work and advancements for the deaf and dumb community.

Keywords : Speech To Text, Text To Speech technique, deaf and dumb community

I. INTRODUCTION

In a large country like India, 63 million people have speaking and hearing disabilities according to WHO in 2016. Such people converse with each other through sign language. But they find it difficult to converse with people who don't know sign language. This can be due to two reasons: There isn't a well- established sign language system in India and there are not many teachers available to teach sign language. Most of the sign language systems used in the Indian media is based on the British sign language system whereas schools and other educational institutes use the American Sign Language system. This creates a language barrier between the people who use American Sign Language (ASL) and British Sign Language (BSL) and many more. Sign language may be a natural method of communication between traditional and deaf dumb folks. Linguistic communication is usually enthusiastic about hand gesture recognition. A gesture is also outlined as a movement, sometimes of hand that expresses a thought. Linguistic communication may be an outlined method of conveyance within which each word or alphabet is delegated some gesture. It's generally difficult for traditional folks to acknowledge the signs properly and perceive what they require to mention.

A gesture-based communication is a language which predominantly utilizes hand developments and body language to impart importance and thought. This can include at the same time consolidating arms or body, and outward appearances to smoothly communicate a speaker's contemplations. The gesture-based communication and the communicated in language have There are similitudes between them, yet there are additionally a few critical contrasts among marked and spoken dialects. Communication via gestures has gotten a typical and wide use on the planet and is at the centers of neighborhood hard of hearing societies.

The improvement of the gadgets that help the hard of hearing and quiet individuals to speak with typical individuals started quite a while prior. They discover troubles to communicate their considerations or to pass on their message to others so that the analysts endeavor various courses so as to deliver a gadget that may give them a superior nature of the life to work in essential circumstances. To accomplish this, the framework joins the utilization of a lot of various modules, for example, motion acknowledgment, communication via gestures investigation and combination, discourse examination and amalgamation, haptic, into a creative multimodal interface accessible to crippled clients. In the late years, there is a fast

increment in the quantity of discourse - crippled casualties because of a few reasons like by birth, oral illnesses, mishaps, and so forth and requirement for the Electronic Assistive.

Not much of thought and consideration is given to address this issue in the society. Many have made an attempt to address this issue using various techniques like capturing image form camera, image to text conversion using Optical Character Recognition (OCR), using Arduino uno boards, Raspberry pi boards, ARM7 processors, Neural Networks.

At present time, smart phones are quite popular and easily accessible. They have features that can help any human being to make a tedious task rather simpler. If person are able to translate voice input in text in real time, it can bring a drastic change in life of deaf and dumb people. At present, technology has enhanced and advanced itself exponentially. Speech recognition is an alternative to keyboard typing. Simply put, you're talking to the machine and the words show on the screen. It has been developed to provide a simple way to write on a computer and can support people with a range of disabilities. It is helpful for users with hand disabilities who often find it though, impossible or painful to type. Gestures may have a limit for use as not all the people are familiar with sign language and also the fact that various region in the world uses different sign languages. Trying to implement a communication mechanism in order to overcome differences between the deaf and dumb community as the normal people. Speech to Text translation allows us to convert perceptible language into text that they can read, edit as well as write on smart phones. Primary concern is to avail a way for deaf and dumb community to get closer to the Technical Era by making use of Speech to Text and Text to Speech technology. The Voice recognition technology is evolving rapidly and is expected to become not only the default input form for smartphones, but also for cars and other home appliances such as TV and fridge. By implementing this technology will be helpful to provide automated voice over for the dumb community. It also provides speech to text translation to interpret the message to the deaf community which in future can be used on internet calls. Text to Speech system can be implemented in various languages. It helps deaf and dumb people to express their feelings and bridge the gap between normal and deaf as well as dumb community. One method to implement Speech to text translation in real time is by making use of HMM i.e. Hidden Markov Model. For text to speech translation, Deep Neural Network can be used. Text to Speech mechanism is implemented in five steps. That is segmentation model that is used to locate boundaries for phonemes, conversion model from grapheme to phoneme, prediction model for phoneme duration and the audio synthetisation model.

II. Literature Survey

1. "Smart Assistant for Deaf and Dumb Using Flexible Resistive Sensor" - K. Kumuda, Preethi k Mane Published 2020

The proposed approach is based on detection of the finger movements and hand gestures to identify the gesture using signal processing kit in LabVIEW software and a data acquisition device (NI USB 6008

DAQ card). The signal is processed and is used to identify the sign shown and concatenate the letters into suitable word and also present the word in audio format. In our work we aim to provide a solution for the trainers in deaf and dumb schools to display letters, with its associated visual display, along with audio output, making learning very interactive and encouraging. It can be used to help treat patients suffering from speech-loss and partial paralysis patient with minimal movement of their fingers. In this project we have implemented the code in LabVIEW platform for twenty-six letters and concatenation of letters up to 6 letters according to American Sign Language. Algorithm can be customized to build the required word length.

Considering the above-mentioned problems existing in the present scenario, we propose a system that uses Smart gloves for translation based on the American Sign Language system. Here, we use flex sensor to obtain the signals from finger based on hand gesture using LabVIEW software. This system is efficient and extremely cost effective. The letters shown by the user is concatenated into words and are represented in both audio and visual output. These smart gloves system is mainly focused to be used as a teaching aid by teachers to introduce and teach sign language for children .Around 60 to 80% of individuals with amputation experience phantom sensations in their amputated limb [3], and the majority of the sensations are painful. These Smart gloves can be used to diagnose phantom limb by using a method similar to the mirror box treatment method. It can also help patients suffering from speech-loss and partial paralysis but are able to move their fingers.

2. “Artificial Intelligence-based Voice Assistant”-Subhash S, Siddesh S, Prajwal N Srivatsa, Ullas A and Santhosh B Published 2020.

The voice assistant is commonly being used in smartphones and laptops. Artificial Intelligent-based Voice assistants are the operating systems that can recognize human voice and respond via integrated voices. This voice assistant will gather the audio from the microphone and then convert that into text, later it is sent through Google Text to Speech. Google Text to Speech engine will convert text into audio file in English language, then that audio is played using play sound package of python programming Language.

The project will give a fair knowledge about the intelligent assistant which is capable of understanding the commands given by the user. Our assistant can easily understand the commands given by the user through vocal media and responds as required. Our assistant performs the most frequently asked requests from the user and makes their task easier. Our voice assistant listens to the command given by the user through the microphone. After listening it will say “don’t listening” and display what the user said and acts accordingly. Voice assistants are all written in programming languages, which listens the verbal commands and respond according to the user’s requests.in this project we have used Python Programming language to build the AI-based Voice assistant.

Advantages

• Communication will be ease and accurate.

• Helps to interpret exact sign to communicate.

Disadvantages

• Slow communication.

• Need to make exact sign for proper communication.

• Communication is dependent on camera response and visibility of sign.

3. “Glove Based Sign Language Interpreter for Deaf and Aphonic Peoples”-G R Meghana, Harshitha H S , L Shyla ,Madan Kumar S M-Published 2020.

This project aims to facilitate individuals by means of a glove based deaf-mute communication interpreter system. The glove is internally equipped with accelerometer. These sensors sense the movement of hands and fingers. For every specific gesture, the accelerometer measures the orientation of hand. The process of these hand gestures is interpreted in controller. Using this device deaf-mute and a standard person can be able to communicate with each other in an affordable and convenient way. This project analyses the data from an instrumented data glove for use in recognition of signs and gestures. A system is developed for recognizing these signs and their conversion into speech. The hardware components used in this project are Renesas microcontroller, ADXL337 accelerometer, voice module, LCD, speaker, regulated power supply.

The method of communication between deaf-mute and normal persons has becomes a difficult task. People who are audio and speech impaired face a lot of difficulty in communicating with others and expressing themselves. Hence, they adopt the method of hand gestures or sign language in order to talk or communicate. But for the normal people find it difficult to comprehend as they are unfamiliar with these signs and gestures, hence resulting in a communication gap between the impaired and the normal ones. The Glove Based Sign Language Interpreter forms a bridge of communication for deaf-mute and normal people. It mainly targets the dumb and deaf to be able to communicate their ideas through hand recognition and voice recognition gesture. Sign language interpreter consisting of accelerometer sensor which helps to measure the movement of the three-axis direction (x,y and z). The accelerometer sensors are placed on the glove. The data from these sensors are sent to the microcontroller for further processing purpose. Once data is recognized at the micro-controller that is sent to the android phone via Bluetooth module. At android phone side, app is developed which is text to speech converter that helps to convert text signal to voice signal. LCD module is also used to display recognized data from the microcontroller.

Advantages

• Fast and easy Communication using action to interpret into voice.

Disadvantages

• Limited word communication.

• Communication depends on sensor sensitivity.

4. "Multipurpose Smart Glove for Deaf and Dumb People"-Pravin Bhalghare ,Vaibhav Chafle,Ameya Bhivgade , Vaibhav Deokar Published 2020.

This system has main purpose to reduce the communication gap between two communities. The main aim of our proposed project is to developed the cost effective system where disable people can communicate with normal people by using hand glove. This means that communication is not barrier between two communities by using smart glove. So disable can also able to grow in their respective field.

This model is a desirable Interpreter which translates. Natural English Sentences as, an text input by Normal Person for Deaf Person and Sign Language, in form of Gesture by a Dumb Person to Synthesized English Words which have a corresponding meaning in Sign Language which interprets a particular thing, as an Audio Output for Normal Person.

Proposed system primarily consist of two sections:

1. Transmitter section

2. Receiver section

Here transmitter section consist smart glove which contain the four flex sensors each on the four fingers of hand namely index, middle, ring and little. The flex sensors give their outputin the form of change in resistance according to their bending angle. The output from the flex sensors is given to the in build Analog to Digital Converter channels of the microcontroller PIC12F683. The processed Analog to Digital Converter values from the microcontroller are compared with the threshold values of each flex sensor for the recognition of a particular gesture. The particular gesture is recognized at the receiving section converted digital data is given to the microcontroller which transmits them through the Radio Frequency module in a serial manner for each value received at Radio Frequency receiver, the microcontroller 18F45K20 gives corresponding commands to the Liquid Crystal Display and the Voice Module.

Advantages

• Communication will be ease and accurate.

• Helps to interpret exact sign to communicate.

Disadvantages

- Slow communication
- Need to make exact sign for proper communication
- Communication is dependent on camera response and visibility of sign.

III. SYSTEM ANALYSIS

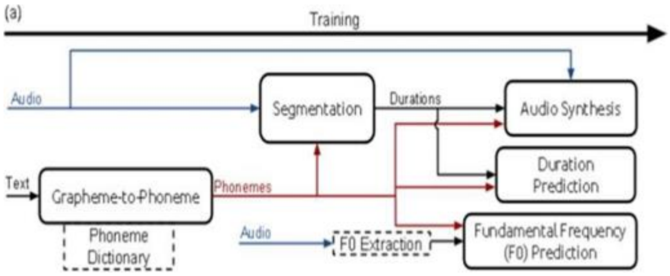
Artificial production of human speech is known as speech synthesis. This machine learning-based technique is applicable in text-to-speech, music generation, speech generation, speech-enabled devices, navigation systems, and accessibility for visually-impaired people.

The two primary technologies in synthetic speech generation are concatenation synthesis and formant/ articulatory synthesis. In that we are using Concatenative synthesis uses high-quality audio clips of different lengths, combining them to form a speech. These audio units must be pre-recorded from a single speaker before the synthesis occurs.

The most important aspect of concatenative synthesis is to find the correct unit length. Depending on the type of units used for concatenation, there are mainly three types of concatenative synthesis, namely,

(1) Domain-specific synthesis
(2) Diphone synthesis
(3) Unit selection synthesis

One of the most commonly used units in the unit selection synthesis is phoneme. A phoneme is one of the units of sound that distinguish one word from another in a particular language. For example, the widely known set of phonetic transcription codes ARPAbet comprises 39 phonemes (not counting lexical stress) of General American English.



The grapheme-to-phoneme model- Initially, it converts the written form of text (i.e. English characters) to voice segments that are converted into coded form with help of a phonemic dictionary.

2. The segmentation model- This model draws boundaries for phonemes from the data input. It illustrates the beginning and end of Audio input provided the voice dataset and a phoneme-by-phoneme description of the input files.

3. The phoneme duration model -This duration model basically predicts the temporal duration of each phoneme in an utterance (i.e phoneme sequence).

4. The fundamental frequency model – This model prognosis if a given phoneme is uttered or not. If it is so, the model is used for forecast of the overall fundamental frequency (F0) in the phoneme's time interval.

5. The audio synthesis model- It collaborates the outcomes of all the previous models and synthesizes the transcripted phoneme sets with respect to the targeted text.

The second phase of text to speech system is inference. In this phase, any of two methodologies is used to take the text as input. Either, text is taken via the grapheme-to-phoneme module or a phoneme dictionary is used to generate phonemes. After this, the phonemes are generated to be given as input to other modules, that are the phoneme duration prediction model and frequency prediction model. These models are used to allocate time intervals of its existence to each phoneme and to forecast a frequency con-tour. Now at the end, local conditioning input features in form of the phonemes, phoneme intervals, and their corresponding frequencies are provided as input to the audio synthesis model.

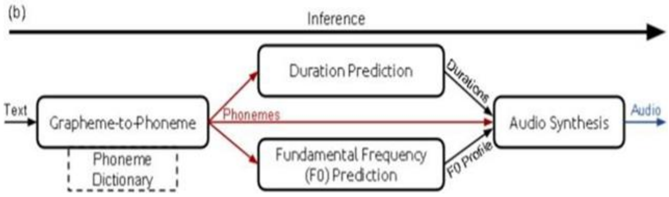


Fig: Inference mechanism

In the above figures 3.1 and 3.2 where input sets are provided on LHS and output sets are provided on RHS. In the given module, time interval i.e. duration as well as frequency is predicted by same Neural Net prepared with a joint loss. Non-learned components are represented by Dotted lines.

It creates the final voice using the phoneme and its attributes. During inference phase all other modules are used but segmentation model. Instead, segmentation model is used to map the voice data set with the phoneme boundaries. These limits are described by the time intervals which are taken reference of, while tutoring the phoneme duration model in first phase. Training of last pass of inference phase is done by the audio marginalized by phoneme set, its interval and F0.

Speech to Text

Conversion (STT) system is widely used in many application areas. In the educational field, STT or speech recognition system is the most effective on deaf or dumb students. The recognition of speech is one the most challenges in speech processing. Speech Recognition can be defined as the process of converting speech signal to a sequence of words by means of Algorithm implemented as a computer program.

With advanced technology, Speech to text translation is much easier than it was few decades or years back. At present time, Speech to text can compile most of the spoken words exactly and translate over 92+x % of total phoneme in multiple languages. Although, this much accuracy is not enough, as 95+x % is required to process a text and transmit it over network for it to be understandable (Stinson et al. 1999: accuracy). To achieve even this 92+x% precision in Automated speech recognition. It refers to a technology that converts spoken words into written text. This technology allows computers to identify and process the words a person speaks into an input device or microphone connected to a computer. Automated speech recognition is independent transcription software designed to convert the spoken language into readable text. It is of two types which are as follows;

1. Direct dialogue conversations: It is a basic version of Automated speech recognition. It consists of machine interface which connects with people. You are required to verbally interact with the computer; the machine tells you to respond with a specific word from a list of words and accordingly, provides response or answer to your request. Automated telephone banking uses this technology to enable customers perform a wide range of financial transactions over the telephone.

2. Natural language conversation: It is a more advanced and sophisticated version of Automated speech recognition. It understands the user'[s speech or written material and responds to the user on the basis of understood content. It enables people to interact with computer using everyday language.

The basic sequence of events that exists in ASR is as follows:

1) A person speaks to the software using an input device like a microphone.
2) The input device creates a wave file of your words.
3) The volume of wave file is normalized and background noises are removed.
4) The cleaned wave file is broken down into phonemes which are the smallest units of sound. There are around 44 phonemes in English.
5) The ASR software analyzes the phonemes, starting from the first phoneme. It uses statistical probability analysis to figure out whole words before making a complete sentence.

6) Now, after understanding the words, the ASR responds in a meaningful way.

The trainer has to fore mostly tutorials the modules with given speech samples, which is a tedious task itself. Though, achieving 100% precision is quite difficult as there are continental accents,that are normally very badly recognized in spite of experiencing intensive training by the module. Noise in environment adds difficulty level in recognition process. Mapping of physical parameters with original voice input is another factor for Automated speech recognition. It is created by the generic module of linguistic as well as the phoneme and the data set from each training and testing result. On the basis of difference of individual physical parameters with results of the training parts, we can conclude the error of Speech To Text implementation. One more thing to be taken note of, if environmental disturbance makes the signal-to- noise-ratio reduce, precision maybe effected negatively by 10+x%. Hidden Markov Model is used for Isolated Word recognition. There are various methods of feature extractions. In recent researches, many feature extraction techniques are commonly used such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Linear Predictive Coding (LPC), Cepstral Analysis and Mel-Frequency Cepstral Coefficient (MFCCs), Kernal based feature extraction based approach, Wavelet Transform and spectral subtraction .Here extracted Mel-Frequency Cepstral Coefficient features for a particular wave file of the utterance 'one.wav'. It has nearly 30 frames where each frame contains 13 coefficients. The phonetic representation of one=w ah n. Initial assumption is that each phoneme to be represented by 3 states S1, S2 and S3.

The first 10 frames (phoneme 'w') will be modelled using 3 states, next 10 frames (phoneme 'ah') using other 3 states and next 10 frames (phoneme 'n') using other 3 states. Thereby each phoneme is represented by 3 states.

• Each phone (consecutive 3 states) is represented by 3*3 transition probability matrix.

• Since there are 3 states, each state is assumed to be represented by single GMM with 4 gaussians. Hence each state will contain Mean vector [4*13], Variance vector [4*13] and priori weights [4*1].

• Each phone is represented by 3 different mean vectors (m1, m2 and m3) where each one each of size 4*13, variance vector (4*13) and Weight vector (4*1). Each phone is represented by a single State transition matrix (3*3) and Emission probability matrix (3*10), 10-frames.

After initialization of Hidden Markov Model parameters, we need to compute the forward probabilities (alpha's) and backward probabilities (beta's) for every phoneme. Further, the state transition matrix and Emission probability matrix for each of the phoneme Hidden Markov Model is updated until there are no significant changes in the log likelihood after successive iterations. After completion of training phase, Viterbi decoding is used to find the hidden state sequence given the observation sequences and the model parameters.

Initially, we need to correctly segment the utterance signal, phonemes don’t have the same length.

Each ith phoneme is described by definite number of frame Ni. (one=30 frame, N1=12, N2=16, N3=8 for example; although there is some overlapping between phonemes). After segmentation, each phoneme is modelled as triphone since it is not possible to avoid the overlapping between adjacent phonemes, so for the given example of one = w ah n, the model will be one= sil-w-ah, w-ah-n, ah-n-sil. Therefore, a phoneme is again sub-segmented into three phones: beginning, centre, end. Each phone (sub-phoneme) is modelled by a GMM associated with an Hidden Markov Model state (in sil-w-ah for instance, S1/GMM1 will model sil-, S2/GMM2 will model -w- and S3/GMM3 will model -ah). A state has two transition probabilities and one emission probability. Addition of an input and output states is provisional. (if N1=12 for the phoneme sil-w-ah, sub-phoneme length maybe N11=2, N12=8, N13=3 for example, overlapping is again unavoidable). To model a phoneme, we need more than one utterance. In case, there are 50 utterances of the same word, only thing need to be done is to segment them correctly into phonemes and then sub- phonemes (the length of the same phoneme, and therefore sub-phoneme, may vary from one utterance to another). For each sub-phoneme collect frames that are associated to it, model them as a Gaussian Mixture Model, and then train the Hidden Markov Model. Gaussian Mixture Model can be shared between Hidden Markov Models.

Algorithms

The feature extraction process is implemented using Mel Frequency Cepstral Coefficients (MFCC) in which speech features are extracted for all the speech samples. Then all these features are given to pattern trainer for training and are trained by Hidden Markov Model to create HMM model for each word. Then viterbi decoding will be used to select the one with maximum likelihood which is nothing but recognized word.

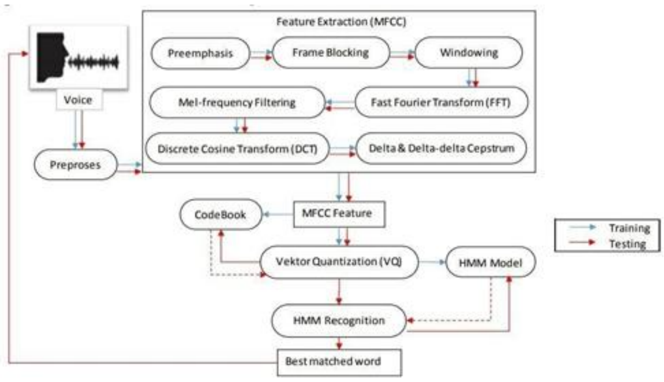
Hidden Markov Model

Hidden Markov Models (HMMs) are widely used probabilistic models, particularly for annotating sequential data with an underlying hidden structure. Patterns in the annotation are often more relevant to study than the hidden structure itself. A typical Hidden Markov Model analysis consists of annotating the observed data using a decoding algorithm and analyzing the annotation to study patterns of interest.

Hidden Markov Model creates stochastic models from known utterances and compares the probability that the unknow utterance was generated by each model. This uses theory from statistics in order to (sort of) arrange our feature vectors into a Markov matrix (chains) that stores probabilities of state transitions. That is, if each of our codewords were to represent some state, the HMM would follow the sequence of state changes and build a model that includes the probabilities of each state progressing to another state.

HMMs are more popular because they can be trained automatically and are simple and computationally Feasible to use HMM considers the speech signal as quasi- static for short durations and models these frames for recognition. It breaks the feature vector of the signal into a number of states and finds the probability of a signal to transit from one state to another. HMMs are simple networks that can generate speech (sequences of cepstral vectors) using a number of states for each model and modeling the short-term spectra associated with each state with, usually, mixtures of multivariate Gaussian distributions (the state output distributions). The parameters of the model are the state transition probabilities and the means, variances and mixture weights that characterize the state output distributions [10]. This uses theory from statistics in order to (sort of) arrange our feature vectors into a Markov matrix (chains) that stores probabilities of state transitions. That is, if each of our code words were to represent some state, the HMM would follow the sequence of state changes and build a model that includes the probabilities of each state progressing to another state.

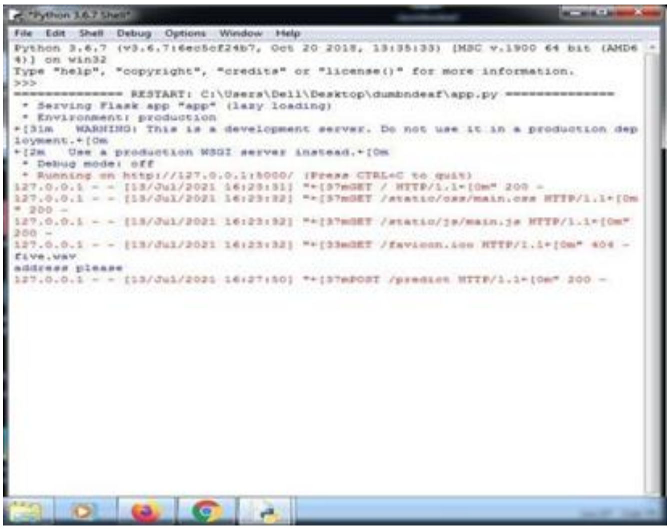
A voice recognition system consists of two phases, namely training phase and verification phase. In the training phase, the voice of the speaker will be recorded and then processed to produce a model form in the database.



IV. SYSTEM ARCHITECTURE

The system created in this study is divided into three parts, namely feature extraction, training, and sound pattern recognition. The following research framework can be seen in Figure 1, each of which will be explained further in the block diagram.

The e-learning section consists of the learning process to recognize objects and the process of evaluating learning outcomes in the form of guessing objects.



V. RESULTS

Fig : Obtained URL which need to be opened in the browser.

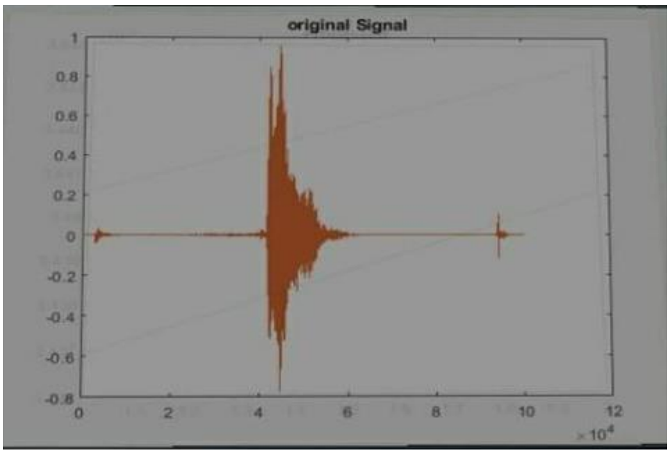


Fig: Reading the uploaded Murmured audio file

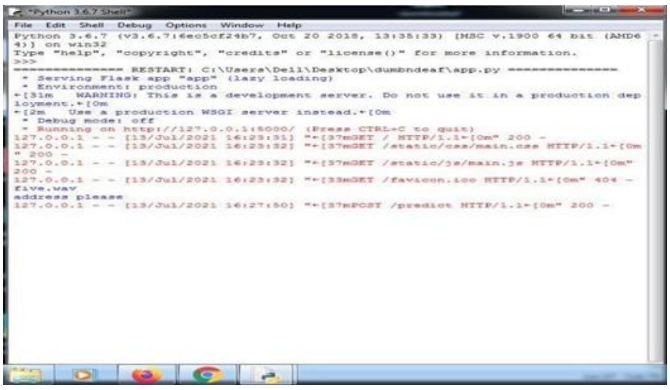


Fig: Obtaining the graph for uploaded audio file.

VI. CONCLUSION

This project implements TTS using deep Neural Network and STT using Hidden Markov Model. It provides a module that can be used for real time conversion which can be used to assist deaf and dumb people to transmit their messages. This system is proposed to improve lifestyle of dumb/ deaf person's. This project is also favorable for degrading the communication difference between the blind person and the dumb person. All over the project is effective and efficient because it is using the TTS modules and further building efficient system for delivery of emotional prosody. This paper is helpful for the industry of people working in the area of designing systems based on Speech synthesis.

VII. CONCLUSION

This project implements TTS using deep Neural Network and STT using Hidden Markov Model. It provides a module that can be used for real time conversion which can be used to assist deaf and dumb people to transmit their messages. This system is proposed to improve lifestyle of dumb/ deaf person's. This project is also favorable for degrading the communication difference between the blind person and the dumb person. All over the project is effective and efficient because it is using the TTS modules and further building efficient system for delivery of emotional prosody. This paper is helpful for the industry of people working in the area of designing systems based on Speech synthesis.

VIII. REFERENCES

[1]. P. Pujol, S. Pol, C. Nadeu, A. Hagen, and H. Bourlard, "Comparison and combination of features in a hybrid HMM/MLP and a HMM/GMM speech recognition system," IEEE Transactions on Speech andAudio processing, 2005 vol. 13, pp. 14-22,.

[2]. E. Zarrouk, Y. B. Ayed, and F. Gargouri, "Hybrid continuous speech recognition systems by HMM, MLP and SVM: a comparative study," International Journal of Speech Technology, vol. 17, pp. 223- 233, 2014.

[3]. M. L. Seltzer, D. Yu, and Y. Wang, "An investigation of deep neural networks for noise robust speech recognition," in proceedings of IEEE International Conference on, 2013, pp. 7398-7402.

[4]. O. Abdel-Hamid, L. Deng, and D. Yu, "Exploring convolutional neural network structures and optimization techniques for speech recognition," In Proceeding of IEEE paper Interspeech, 2013, pp. 3366- 3370.

[5]. Y. Zheng, "Acoustic modeling and feature selection for speech recognition," In Proceeding of IEEE, 2005, pp.345-350.

[6]. Deepa V.Jose, Alfateh Mustafa, Sharan R,"A Novel Model for Speech to Text Conversion" International Refereed Journal of Engineering and Science (IRJES)ISSN (Online), Volume 3, Issue 1 2319-2327,2014

[7]. Chen, Jingdong, Chrzanowski, Mike, Coates, Adam, Diamos, Greg, et al. Deep speech 2"End-to- end speechrecognition in English and mandarin", arXiv preprint arXiv,6th edition, 2015.

[8]. KyungHyun, and Bengio, Yoshua " Empirical evaluation of gated re-current neural networks on sequencemodelling", arXivpreprint arXiv,8th edition ,2014.

[9]. Saadman Shahid Chowdury, Atiar Talukdar, Ashik Mahmud, Tanzilur Rahman"Domain specific Intelligent personal assistant with bilingual voice command processing",In proceeding of IEEE conference, 2018,pp.117-121.

[10]. Veton Kepuska and Gamal Bohota,"Next generation of virtual assistant (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)",In Proceeding of IEEE conference, 2018,pp.234-240.

[11]. Laura BURbach, Patrick Halbach, Nils Plettenberg, Johannes Nakyama, Matrina Ziefle, Andre Calero Valdez"Ok google, Hey Siri, Alexa. Acceptance relevant of virtual voice assistants",In Proceeding of International communication conference IEEE, 2019,pp.567-571.

Cite this article as :

Suchethana H C, Arun Kumar P, "Virtual talk assistance for Dumb and Deaf", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 3, pp. 96-105, May-June 2022. Available at doi : https://doi.org/10.32628/IJSRSET122936
Journal URL : https://ijsrset.com/IJSRSET122936