MANTECH Publications
Journal of Digital Image Processing and Computer Vision
Volume 2 Issue 1

Formal Malayalam Speech to Text Converter

Tany Thomas, Ammu Appukuttan, Haritha Melvin, Linda Sara Mathew, Chaitanya C
Department of Computer Science and Engineering
M.A. College of Engineering, Kothamangalam
Corresponding Author: *tanythomas95@gmail.com*

Abstract

Automatic Speech Recognition has been a subject of intensive research for the last many decades. It is the process of converting a speech signal to a set of words, by means of an algorithm implemented as a computer program. The goal of this project is to develop a Formal Malyalam Speech to Text Converter for the Malayalam language. The main components of Automatic Speech Recognition system are feature extraction, training and recognition. Feed forwarding technique is used as a front-end to extract acoustic features from the input signal. Resilient Propagation Algorithm is used to train the Artificial Neural Network (ANN) for classification and recognition purpose. The developed system is integrated to open Office Writer. Different applications include malayalam voice search, user friendly editor for visually challenged people, and so on.

Keywords: *Feed forwarding, Resilient Propagation, Artificial Neural Network*

I. INTRODUCTION

In modern civilized societies speech is one of the common methods for communication. Different ideas formed in the mind of the speaker are communicated by speech in the form of words, phrases, and sentences by applying some proper grammatical rules. Speech is the primary mode of communication among human beings and also the most natural and efficient form of exchanging information. Language technologies can provide solutions in the form of ordinary interfaces so the digital content can reach to the masses and facilitate the exchange of information across different people speaking different

languages. These technologies play a vital role in multi-lingual societies such as India which has about 1652 dialects/native languages. In this paper, we propose a Formal Malayalam Speech to Text Converter(FMSTC) which converts formal malayalam speech to corresponding malayalam text. FMSTC takes input from the microphone in the form of speech & then it is converted to text which is displayed on the screen.

Speech processing is the study of speech signals, and the various methods which are used to process them. In this process various applications such as speech coding, speech synthesis, speech recognition and speaker recognition technologies; speech processing is employed. Among the above, speech recognition is the most important one. The main purpose of speech recognition is to convert the acoustic signal obtained from a microphone or a telephone to generate a set of words. Automatic Speech Recognition [3] has been a subject of intensive research for the last many decades .It is the process of converting a speech signal to a set of words, by means of an algorithm implemented as a computer program. Theoretically, it should be possible to recognize speech directly from the digitized waveform. Some form of feature extraction should be used to reduce the large variability of the speech signal. Encog framework[2] is used to implement the neural network in this paper.

II. EXISTING SYSTEM

Accurate and computationally efficient means of recognizing continuous speech has been a subject of research in recent years. The developments are done by improving technologies, computer systems and communication ways. These parallel developments led the way to the applications we use today for converting speech into text. Currently, two major options are available for providing real-time speech-to-text services: Computer assisted note taking (CAN), Communication access (or computer aided) real-time translation. There is a lot of difference between these two methods in their process of generating speech to text in real time, with respect to the circumstances under which the methods can be properly used and with respect to the amount of training which is will help to convert speech to text successfully.

Small vocabulary, speaker independent continuous Malayalam speech recognition systems can be based on Hidden Markov

Models (HMMs)[4]. Continuous density HMM, which is used in this work to model phonemes, represents the general case where the observation probability density functions (pdfs) are continuous. The observation pdf is approximated using a Gaussian mixture density. Mel-frequency Cepstral Coefficients (MFCC) method is used to extract acoustic features from the input signal. To represent temporal variations in the speech signal, the first and second order derivatives of MFCC are added to the set of static parameters. The training and decoding are performed by the Baum-Welch and Viterbi algorithms respectively.

Wavelet[5] is an alternate to conventional Fast Fourier Transforms (FFT). Feature extraction involves information retrieval from the audio signal.

III. FORMAL MALAYALAM SPEECH TO TEXT CONVERTOR

FMSTC converts speech to text in malayalam font. Speech is captured using a high quality microphone which is then converted to corresponding English font using an Application Programming Interface. The normalized form of this text is fed as input to the neural network in which training occurs.The corresponding malayalam text is displayed as the final result.

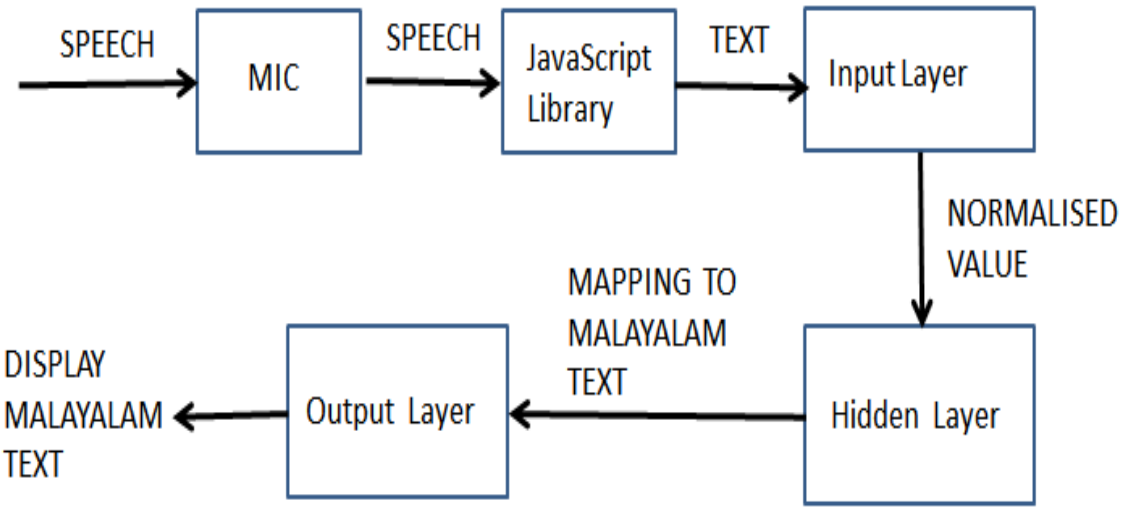

Figure 3.1Basic architecture of FMSTC

IV. SYSTEM DESIGN AND IMPLEMENTATION

A. Automatic Speech Recognition

The speech from the user captured by the microphone is fed to a speech API. It processes the speech, performs feature extraction and yields the corresponding English text. This English text is displayed on the screen. Streaming recognition recognizes live audio as it is captured from a microphone or other audio source. The audio stream is split into frames and sent in consecutive Streaming Recognize Request messages. Any frame size is acceptable.

B. Input Layer

The English text is sent to input layer of the neural network where it is checked if this word is already present in the training set.If yes, it is sent for further processing. If not, an error message is displayed on the screen. If it is present, each character in the english-converted word and its corresponding malayalam word in the training set is converted to its corresponding double value. This set of double values are further normalised using normalisation helper classes to obtain a value within the range (-1 to 1).These normalised values are passed on to the next layer for further processing.

C. Hidden layer

In this layer, Resilient Propagation Algorithm[1] is used to train the Artificial Neural Network (ANN) for classification and recognition purpose. Here, the mathematical representation of speech is mapped to malayalam font. The normalised values are fed to the hidden layer. The normalized values of the english word and the corresponding malayalam word are compared, the weights are adjusted repeatedly to achieve a minimal error percentage.

D. Output layer

After achieving the optimal error percentage, the corresponding malayalam text is displayed on the screen.

V. APPLICATIONS OF FMSTC

The application field of Speech to Text is expanding fast whilst its quality is also increasing steadily. Speech synthesis systems are also becoming more affordable for common customers, which makes these systems more suitable for everyday use & becomes a cost effective. Most of the research work done in this area is in English, Arabic and Mandarin. There is minimal work done in native Indian languages. This paper focuses on

Malayalam speech to text conversion, and is intended to be used by Malayalees who do not know English. FMSTC can be used in applications such as:

- For native Malayalam speakers to use search engines which supports voice search
- User friendly editor for visually challenged people
- Interactive voice response system (IVRS)
- Automated teller machines (ATMs) Data entry work
- In classroom works for disabled students

CONCLUSION

Speech recognition technology has made a remarkable progress in the past 5 - 10 years. Based on the progress, various application systems have been developed using dictation and spoken dialogue technology.This paper has illustrated recognition system for Malayalam language words using Resilient Propagation Algorithm.It recognizes any combination of formal Malayalam words pronounced with pause between the words.The proposed system provides a user-friendly interface for the conversion. The accuracy of the system was found to be satisfactory.The accuracy can be further improved by using larger training data; including utterance from a large of speakers with variations in age and accent.

REFERENCES

[1] Fábio M. Soares,Alan M.F. Souza,Neural Network Programming with Java,1st ed, Packt Publishing Ltd.,2016,pp.33-108

[2] Jeff Heaton, Programming Neural Networks with Encog3 in Java,1st ed,Heaton Research, Inc,2011,pp.52-153

[3] Nuzhat Atiqua Nafis,Md. Safaet Hossain," Speech to Text Conversion in Real-time", International Journal of Innovation and Scientific Research, Vol. 17 No. 2 ,pp. 271-277, Aug. 2015

[4] Steve young , "A review of large-vocabulary continuous-speech recognition", IEEE signal processing Magazine ,September 1996

[5] Vimal Krishnan,V R Athulya Jayakumar,Babu Anto.P," Speech Recognition of Isolated Malayalam Words

Using Wavelet Features and Artificial Neural Network",4th IEEE International Symposium on Electronic Design, Test & Applications,October.2008

[6] T M Thasleema,V Kabeer,N K Narayanan," Malayalam Vowel Recognition Based On Linear Predictive Coding Parameters and k-NN Algorithm", International Conference on Computational Intelligence and Multimedia Applications,2007

[7] L.Rabiner, B. Juang, "Fundamentals of Speech Recognition", Pearson Education, 1993, pp 20-30