International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

MALAYALAM SPEECH TO TEXT CONVERSION

Anandhu Sasikuttan[1], Ashish James[1], Ajay P Mathews[1], Abhishek M.P[1], Prof. Kishore Sebastian[2]

Department of Computer Science and Engineering, St Joseph's College of Engineering and Technology Palai, Kerala, India

---

Abstract - *The Speech recognition for many languages is getting more common in the present scenario. In Malayalam language, recognizing speech is a very difficult task. The goal of this project is to develop a Formal Malayalam Speech to Text Converter for the Malayalam language. The uttered word which is the input for the system and displayed the corresponding Malayalam as the output. In this project we are using data analysis technique and artificial intelligence for Malayalam speech to Malayalam text conversion. In data analysis using text analysis technique for the identification of Malayalam words. The main purpose of data analysis is to extract useful information from dataset and taking decisions based upon the data analysis.*

1. 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. The most common method to interact with computers is using a keyboard and mouse. When a need to enter a large amount of data to a computer that process will be difficult and time consuming. In order to solve the problem the mode of operation should be changed. Speech is the primary mode of communication among human beings and also the most natural and efficient way of exchanging information. If a system can understand what a human speaks, then it will be the best form of interaction between a human and a computer. Speech to text systems take speech as input, recognize it and convert it into text. Speech to text system support many applications such as, an aid for illiterate people, telephone directory assistance, in hospitals for health care instruments, in banking, in mobile phones etc. The proposed system converts Malayalam speech into a text form. Malayalam is one of the Dravidian family languages and is the official language spoken by Keralites. In this model Speech to Text conversion process takes place by Data Analysis technique and Artificial Intelligence. Data Analysis is used to extract useful information from a dataset and to make decisions based upon the data analysis. The voice identification of the person is done by speaking the words via a microphone in a noiseless background to achieve higher efficiency to the system. The words in the voice are extracted by using some Artificial Intelligence technique such as NLTK for the conversion of words and a Word Tokenizer for identifying words in the voice. The final text is obtained by identifying the Malayalam words from the extracted words by using Data Analysis.

2. EXISTING SYSTEMS

Some similar applications or web pages that are useful for flood victims or trapped ones are listed below:

- **Hidden Markov Model (HMM)**

HMM is a statistical model used in speech recognition as the speech signal can be viewed as a piece wise stationary signal or a short-time stationary signal [4]. HMM, models are utilizable for genuine-time verbalization to text conversion for mobile users. They allow us to predict the unknown variables based on the set of observed variables [5]. We know that the speech is a random process and the purpose of the HMM is to find the parameters of some random process in a well-defined manner

- **Wavelet**

Wavelet analysis is used to provide a more efficient way of representation of the speech signal[6]. It is considered to be an effective and powerful method for extracting the features for speech processing. The procedure of the wavelet analysis is to adopt a wavelet prototype function called mother wavelet. Any signal of speech is to be delivered translated versions and measured by the mother wavelet. Based on the time-frequency multi-resolution property of wavelet transform, the input speech signal is decomposed into various frequency channels. The wavelet have the capacity to distinguish between different properties of high frequency low amplitude spectral components and low frequency large amplitude spectral components. In addition to that the neural network classifier improves the recognition performance of the system up to an extent.

- **DTW**

Dynamic time warping is a technique used for locating similarity between two sequences that will vary in time or speed [1]. DTW was successfully employed to match the test speech with the available reference patterns. DTW is an algorithm for measuring similarity between two sequences which can vary in time or speed. The sequences are warped nonlinearly within the time dimension to see a measure of their similarity independent of certain nonlinear variations within the time dimension [2]. DTW was recognized as the

most fitted method for speech recognition thanks to its capability to address different speaking speeds.

- **Artificial Neural Network Classier (ANN) based Cuckoo Search Optimization**

With Cuckoo Search Optimization technique is employed for better communication, better recognition and to remove unwanted noise[3]. ASR is built for a more robust interface of human and machine interaction and it's three-step process includes Pre-processing of signals, feature extraction, classification. Pre-processing of the speech signals is the most vital part of speech recognition it's accustomed to remove avoidable waveforms of the signal. The signals are provided to the high-pass filters in order to reduce the background noises. In Cuckoo Search Optimization, ANN is implemented by a two-layer Feed Forward Backpropagation Neural Network with 3 units, two input units, three Hidden units and one output unit. Two inputs are provided to the input layer which have two features extracted which are MFCC and LPCC features. These features are given as input within which networks get trained and it engenders a corresponding output

3. PROPOSED METHOD

The main intention of our project is to encourage the use of our native language and to help illiterate people for the easy typing of the text. The concept involves the recognition of voice through a microphone. The background noise is reduced by using a noise filter cap over the mic. The words are extracted from the input voice by using some artificial intelligence and feature extraction. By using NLTK we convert the word and these words in the uttered speech are identified using word tokenizer. The extracted words are then compared with the pre-trained data set by using data analysis. Finally, the mapping of the words from the voice to Malayalam words in the data set and join that word using the string joining method. Now the corresponding text to the uttered speech displays on the screen.

4. EXPERIMENTAL ANALYSIS

The speech to text conversion of our proposed system consists of two phases; training phase and testing phase. The voice of the speaker that is recorded through a mic is considered as the input to the system.

Training phase:

In this phase we create datasets of voice and their corresponding texts of thousands of Malayalam words. The words in the datasets are some of the most commonly used Malayalam words which help in improving the accuracy of the system. The input voice for the training set is recorded from a noise free environment.

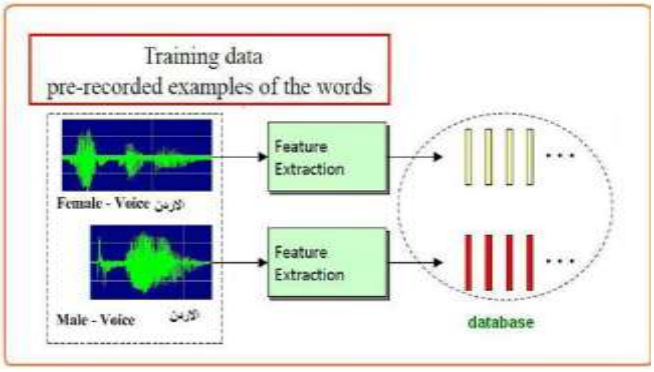

Fig :1 Training Phase

വേണ്ടി	for	vendi
മേല്	on	mellu
ഉണ്ട്	are	und
കൂടെ	with	kooda
അവർ	they	avar
എന്നു	be	ennu
ഭൂതകാല	at	bhoothakalam
ഒന്ന്	one	onu
ഞങ്ങൾക്കുണ്ട്	have	njaglak und
ഈ	this	ee
നിന്ന്	from	nennu
കൊണ്ട്	by	kondu
ചൂടുള്ള	hot	chudulla
വാക്ക്	word	vaku

Fig : 2 Dataset

Testing phase:

In this phase we evaluate the trained datasets and predict the accuracy of the system

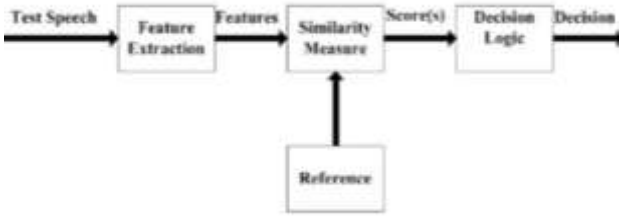

Fig :3 Testing Phase

The system is made familiar with different Malayalam words in order to improve the quality and performance of the proposed method

5. SYSTEM IMPLEMENTATION

5.1 Software Implementation

Anaconda is a free and open-source software that makes Python packages easy to use. It has built in applications like Spyder, Jupyter Notebook, CUDA in which each application takes different tasks. Here we use Jupyter Notebook to write the entire code. Libraries like Speech Recognizer, Pandas, PyAudio etc are used.

- **Speech Recognition**

Speech recognition is a computer software program or hardware device with the ability to decode the human voice. We use python's speech recognizer libraries in our system. The input speech is received through an external mic. The recorded voice is then stored to a variable for further processing

- **PyAudio**

PyAudio provides Python bindings for PortAudio, the cross-platform audio I/O library. The dataset which is the "Data Set. xlxs" file is provided which contains the Malayalam words with their English meaning and the Malayalam word itself written in English and named as Manglish.

- **Pandas**

Pandas is a software library written for the Python programming language for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series.It is used for reading the pre-recorded dataset that is used for training the system.

- **Word Tokenizer**

Word tokenization is the process of splitting a large sample of text into words. This is a requirement in natural language processing tasks where each word needs to be captured. They separate the words from the input speech in their respective order.

- **Mapping**

Mapping is a matching process where the points of one set are matched against the points of another set. The values in the dataset are stored in the form of NumPy which are large python libraries. The tokenized input words are compared with the manglish words of the dataset. The corresponding Malayalam word to the matched manglish word will be the converted text.

5.2 Hardware Implementation

The hardware part of the system consists only of a Microphone. A Microphone is a device that converts sound into an electrical signal. We use a microphone to catch the noise free voice signal to convert into text form. Here we are using a pair of headphones that has a microphone inbuilt to process the speech signal into text format.

Fig:4 A headphone with mic

6. FUTURE SCOPE

The scope of our project Malayalam speech to text conversion is not so vast. Only limited scopes are available for this project. Many other apps and products are available based on malayalam speech to text conversion. Main aim of this project is Documentation purpose. We know that Govt offices use Malayalam scripts for documentation purpose. There is little difficulty in typing Malayalam words through keyboard. Here there is no such difficulties, we just only want to speak what we want to type through this product. And we get the corresponding malayalam word through this product. It also helps the illiterate people to write malayalam words and we promote the use of our native language in our daily life.

RESULT

The Figure 5 shows the login part of the system.

Fig: 5 Step 1

Clicking on the Login button it will take us to the next tab as shown in the figure 6.

Fig: 6 Step 2

After clicking on the Test Now button we can speak through the mic. That will be recognized by the system.

Fig: 7 Step 3

Finally it shows the recognized Malayalam word and is displayed in the tab.

CONCLUSION

Speech recognition technology has made a noticeable development in the past few 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 python's speech recognition libraries. It apperceives any amalgamation of formal Malayalam words pronounced with a gap between the words. The proposed system provides a simple and basic interface for the conversion as it is mainly intended for the illiterate people. The accuracy of the system was found to be satisfactory. The accuracy can be further improved by using larger training data; including the varying accents spread across different parts of Kerala. This will make the system flexible and adaptable to any environment.

REFERENCES

[1] Santosh V. Chapaneri, "Spoken Digits Recognition using Weighted MFCC and Improved Features for Dynamic Time Warping," International Journal of Computer Applications, vol.40, no.3, Feb. 2012.

[2] Shivanker Dev Dhingra, Geeta Nijhawan, Poonam Pandit, "Isolated Speech Recognition Using MFCC and DTW," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol.2, Issue 8, Aug. 2013.

[3] 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

[4] T. Masuko, "HMM-based speech synthesis and its applications," PhD Dissertation, Nov 2002.

[5] J. Lattore, K. Iwano, and S. Furui, "New approach to the polyglot speech generation by means of an HMM-based speaker adaptable synthesizer," Speech Communication, vol. 48, pp. 1227–1242, 2006.

[6] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE transactions on Pattern Analysis Machine Intelligence, Vol. 11 1989, pp.674-693.