

Speech Recognition of Malayalam Numbers using Kaldi

Arunima Prasad J S

Mtech, Dept of Electronics & Communication
Govt. Engineering College Bartonhill, TVPM
Kerala, India

Aparna S Thampy

Assistant Professor, Dept of Electronics & Communication
Govt. Engineering College Bartonhill, TVPM
Kerala, India

Abstract— Speech is a straightforward and usable method of correspondence between people for trading data, however these days people aren't restricted to interfacing yet even to the various machines in our lives, most significant it is a computer. So this correspondence strategy is frequently utilized among PCs and people known as Human Computer Interface (HCI). If a system can understand human language, then it is the best method of interaction between a human and a computer. When a person speak with his own language, the machine must be able to understand it. Natural language processing (NLP) have significant role in these speech Recognition system which converts the speech from human being to text of respective language. Such systems take speech as input, recognize it and convert it into text. Malayalam is the official language of Kerala and it is spoken by 32 million people in India. Malayalam Digit Recognition have significant role in the developing technical field. The system which recognizes the Malayalam numbers spoken and convert it into text. It can be done using Deep learning technique by Kaldi toolkit. The framework utilizes Mel Frequency cepstral Coefficient (MFCC) with cepstral Mean and Variance Normalization (CMVN) and I-Vectors as feature extraction for signal handling . The framework is prepared on a test set of various qualities included dataset.

Keywords—: Kaldi, MFCC, DNN

I. INTRODUCTION

Automatic speech recognition is a highly technical thing that makes machine turn the speech sign to the relating text or order subsequent to perceiving and understanding. ASR consolidates the extraction and confirmation of the acoustic part, the acoustic model, and the language model. The ASR can be act in any customary vernaculars, for instance, Malayalam, Tamil, Telugu, etc. In Malayalam it uses a ton of expansion [1]

Malayalam has a place with the Dravidian group of dialects and is one of the four significant dialects of this family with a rich scholarly custom. Malayalam is the mother tongue of Kerala, one of the southern state of India, and in the association region of Lakshadweep. In Malayalam language, individuals talk in various slangs. Individuals from each piece of Kerala utilize an alternate shoptalk. There are 14 regions in Kerala, and individuals live in each spot talk with various shoptalk. For every well known word individuals from every locale utilize different shoptalk to articulate it or offer something else entirely. If there should be an occurrence of Digit likewise these intricacy shows up. The framework make simple the issue. It might accommodating in numerous specialized field in which numbers can entered by talk. There

are different spoken structures in Malayalam despite the fact that the abstract lingo all through Kerala is practically uniform [3].

II. EXISTING SYSTEM

The are many works reported on Malayalam Speech Recognition. But less work has been reported on Digit Recognition. The existing work only included the single digit values, i.e, From only Zero to Nine. For large application purpose it is not satisfactory.

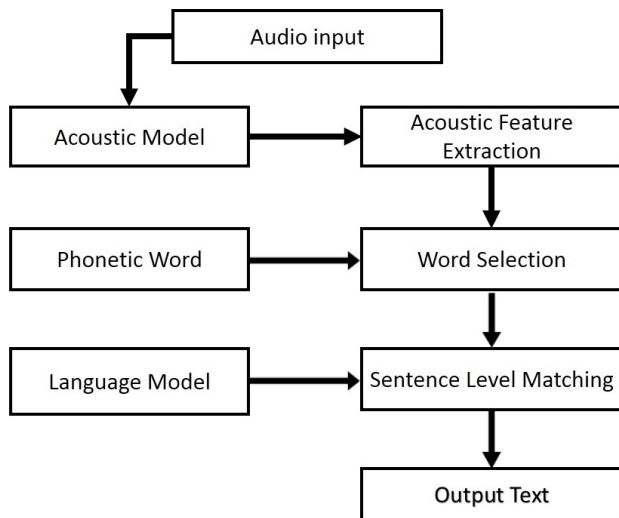
III. MOTIVATION

Malayalam included in the twenty two scheduled languages of India. Government of India declared it as classical language in 2013. Malayalam Digit Recognition have many applications in real life. For example, the number of people who does not use ATM are less, but some people with lack of knowledge of English language and use of computer, cannot use the ATM for their needs. At that time they can use any machine by the very best communication medium, speech. They can access the machine by speaking the PIN amount, etc. With the advent of Malayalam Digit Recognition, physically challenged people will also be able to access machines like ATM's Computers etc. Therefore, designing such a helpful system like Malayalam speech recognition, especially Digit recognition will be helpful for the people and also it will be a great innovation. And also the dataset available for Malayalam Especially for Malayalam Digit is less.

IV. METHODOLOGY

The system is an advanced Speech Recognition System for Malayalam Language which recognize Malayalam Numbers using the speech recognition tool Kaldi and Deep Neural Network (DNN). The input signal is the Dataset of Spoken Malayalam number file. i.e, Collection of recorded voices of Malayalam numbers. The system recognition is carried out by processing the speech dataset by the toolkit Kaldi which is used for testing and training purposes. The Deep Neural Network (DNN) is used for the acoustic modelling. Before the testing and training phase, the database should be feature extracted because of the need of phonetically important characteristics of speech. This can be separated by the feature extraction module which comprises MFCC, CMVN, and I-vectors.

A. Architecture



B. Steps Included

- Process incoming wav speech
- From wave signal, extract acoustic features using acoustic model.
- Linking those features to words or vocabulary or lexicon.
- Language model or grammar defines how words can be connected to each.

IV. KALDI

Kaldi toolkit is an open-source tool stash for speech recognition written in C++ and authorized under the Apache License v2.0. The objective of Kaldi is to have a versatile code that is direct, alter and expand. The instruments total on the for the most part used Unix-like structures and on Microsoft Windows. Kaldi is conveyed under the Apache License v2.0, which is significantly nonrestrictive, making it sensible for a wide neighborhood clients [5].

A. Features

- Reconciliation with Finite State Transducers
- Broad linear algebra support
- BLAS and LAPACK schedule
- Extensible plan: scores.
- Open permit
- Complete recipes
- Careful testing:

B. Parts of Kaldi

- 1) Preprocessing and Feature Extraction
- 2) Modelling
- 3) Training

V. DATA COLLECTION

A. Speech Data Collection

For the preparation of spoken digit data, the speech is recorded in both noise and noiseless environment.

The framework is intended to perceive Malayalam digits, therefore the length or size of the vocabulary is eleven (counting quiet). For noisy and noiseless environment, a top quality receiver with headset can be utilized which includes mouthpiece with 70Hz to 16000 Hz of recurrence range. The recording should be possible with 16 kHz inspecting recurrence quantized by 16 bit. To catch all the acoustic property of the words, the training dataset is intended. A Transcription document is made for every expression of the speaker and a language word reference is made for each word in the string. These are stored in discrete records. For noisy environment, the recording manner is same as that of the recording in the noiseless environment. The recording device used is show in the next section. It is better and advantageous.

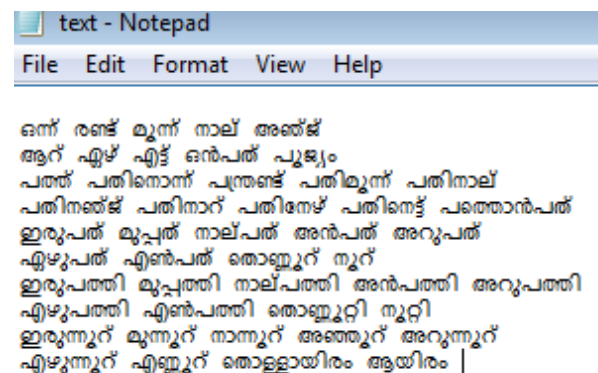
B. Recording

Recording is done in both noisy and noiseless environment. In both cases the recording device Roland R-07 is used because it can capture both noisy and noiseless audios. The numbers from zero to 9999 are recorded with files groups The specifications used for recording are

- Frequency = 48KHz
- Sampling Rate = 16 bit
- Recording of each audio file is saved as .wav format
- The editing includes noise reduction etc. are done by using the Software Audacity.
- Recording is done with 10 different persons with 5 male and 5 female with different age groups and slangs.

C. Text Data Collection

The collection of text data is done by creating a text file with the text of numbers spoken by the speakers. Fig. Shows an example



Each line in the transcript file corresponds to each text file with respect to the recorded audio file.

VI. FEATURE EXTRACTION

A. Mel Frequency Cepstral coefficient (MFCC)

MFCC is a popular technique for extracting features. To catch the phonetically significant attributes of speech, Mel Frequency scale is used to express the signal. This scale has below 1000Hz linearly frequency spacing and a above 1000Hz logarithmic spacing. MFCCs are less involves to the states of being of the speakers' vocal rope, contrasted with the speech wave structures.

B. Cepstral Mean and Variance Normalization (CMVN)

Cepstral mean and fluctuation standardization (CMVN) is a computationally proficient standardization strategy for strong speech acknowledgment. CMVN minimizes noise distortion for robust feature extraction by linearly transforming the cepstral coefficients to have the same segmental statistics

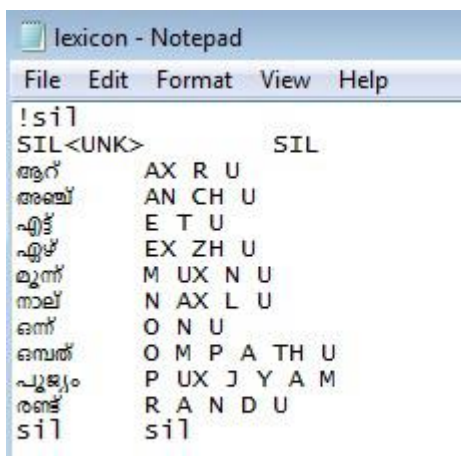
C. I-Vectors

I-Vectors are utilized for addressing the style of every sound expression or speaker. I-Vectors are utilized for better comprehension of the changes inside the kaldi.

VII. EXPERIMENTAL ANALYSIS

A. Creation of Lexicon file

To train the Malayalam Database, a lexicon file is needed. The lexicon file is the phonetic transcription of each number



B. Training of data

The training of data is done by eleven different speakers with both noise and noiseless environment. The Malayalam numbers are from zero to 99999. The overall WER is 35 %

C. Word Error Rate per Speakers

The fig shows the WER per each speaker. The WER can be varies according to the speaker with their sound, voice, slang etc. and also depends on the environment .i.e, noisy or noiseless.

PK	id	#SENT	#WORD	Corr	Sub	Ins	Del	Err	S.Err
ML001	raw	2	14	3	0	0	11	11	2
ML001	sys	2	14	21.43	0.00	0.00	78.57	78.57	100.00
ML002	raw	2	14	0	0	0	14	14	2
ML002	sys	2	14	0.00	0.00	0.00	100.00	100.00	100.00
ML003	raw	2	14	1	2	0	11	15	2
ML003	sys	2	14	7.14	14.29	0.00	78.57	92.86	100.00
ML004	raw	2	6	1	0	0	5	5	2
ML004	sys	2	6	16.67	0.00	0.00	83.33	83.33	100.00
ML005	raw	5	25	20	2	2	3	7	5
ML005	sys	5	25	80.00	8.00	8.00	12.00	28.00	100.00
ML006	raw	3	15	11	3	0	1	4	3
ML006	sys	3	15	73.33	20.00	0.00	6.67	26.67	100.00
ML007	raw	3	15	7	6	1	2	9	3
ML007	sys	3	15	46.67	40.00	6.67	13.33	60.00	100.00
ML008	raw	3	13	10	0	0	3	3	3
ML008	sys	3	13	76.92	0.00	0.00	23.08	23.08	100.00
ML010	raw	3	14	11	0	2	3	5	3
ML010	sys	3	14	78.57	0.00	14.29	21.43	35.71	100.00
ML011	raw	3	15	11	2	0	2	4	3
ML011	sys	3	15	73.33	13.33	0.00	13.33	26.67	100.00
SUM	raw	28	145	75	15	5	55	75	28
SUM	sys	28	145	51.72	10.34	3.45	37.93	51.72	100.00

FIG WER PER SPEAKER

D. Testing of some examples

The fig shows the testing of some example datasets. It shows 75% accuracy

```

Input1 അപരിചിതമായിരിക്കുന്നു എന്നായിരിക്കുമ്പോൾ നാലായിരിക്കുമ്പോൾ നാലായിരിക്കുമ്പോൾ
Input10 ഒട്ടെ അപരിചിതമായിരിക്കുന്നു
Input11 ഒട്ടെ അപരിചിതമായിരിക്കുന്നു
Input12 മുൻ അപരിചിതമായിരിക്കുന്നു
Input13 മുൻ അപരിചിതമായിരിക്കുന്നു
Input14 മുൻ അപരിചിതമായിരിക്കുന്നു
Input15 നാലായിരിക്കുമ്പോൾ മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു
Input2 മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു
Input3 മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു
Input4 മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു
Input5 മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു
Input6 മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു
Input7 മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു
Input8 മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു
Input9 മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു മുൻ അപരിചിതമായിരിക്കുന്നു

```

CONCLUSION

The system is an advanced method for the recognition of Malayalam Numbers using Deep Neural Network in Kaldi. Malayalam Digit Recognition have many applications in real life scenario. The developed system is one of the innovations for the physically challenged, partially blind and people with lack of knowledge of English language and usage of computer. Spoken number Recognition framework gives an easy to understand point of interaction to taking care of numeric information into computers. The precision of the framework is satisfactory

FUTURE SCOPE

The system can be advances to more accurate level by adding more datasets. The WER can be reduces by the method of training. The proposed system can be converted to advanced system which can recognize real time inputs.

REFERENCES

- [1] Kurian, C. and K. Balakrishnan, Speech recognition of malayalam numbers. In 2009
- [2] KURIAN, C. and K. BALAKRISHNAN (2013). Connected digit speech recognition
- [3] P, A. K. A. P. S. A. D. (2021). Malayalam speech recognition. International Journal of Innovative Science and Research Technology, 6, 1323-1325.
- [4] Lekshmi, K. R. and E. Sherly, An asr system for malayalam short stories using deep neural network in kaldi. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). 2021
- [5] Povey1, D. and A. Ghoshal (2020). The kaldi speech recognition toolkit. IEEE, 6, 35-39..

-
- [6] Arun HP, S. R. A. S. A., Jithin Kunjumon (2021). Malayalam speech to text conversion using deep learning. IOSR Journal of Engineering (IOSRJEN), 11, 24–30
- [7] B, B. L., G. Anu, K. R. Sreelakshmi, and L. Mary, Continuous speech recognition system for malayalam language using kaldi. In 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR).2018.
- [8] Moncy, A. M., A. M., H. Jasmin, and R. Rajan, Automatic speech recognition in malayalam using dnn-based acoustic modelling. In 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS). 2020.
- [9] S., R., A. Joseph, and A. B. K.K., Isolated digit recognition for malayalam- an application perspective. In 2013 International Conference on Control Communication and Computing (ICCC). 2013.