Isolated Spoken Word Identification in Malayalam using Mel-frequency Cepstral Coefficients and K-means clustering

Sreejith C [1], Reghuraj P C [2]

[1] M.Tech Computational Linguistics
Department of Computer Science and Engineering
Govt. Engineering College
Sreekrishnapuram, Palakkad, Kerala, India
sreejithc321@gmail.com

[2] Professor
Department of Computer Science and Engineering
Govt. Engineering College
Sreekrishnapuram, Palakkad, Kerala, India
pcreghu@gmail.com

Abstract: *This paper proposes an approach to recognize isolated spoken Malayalam words. The paper deals with a speech feature extraction technique based on MFCC and K-mean clustering. We used six Malayalam words for the experiment and hundred speakers are used to identify these words in the testing phase. The words and stored in a database and later identified. MFCCs are calculated in both training and testing phase for different words, once in a training session and once in a testing session later. The word is identified according to the minimum quantization distance which is calculated between features of each word in training phase and the individual word in testing phase. This system is proposed for real time, speaker- independent word recognition systems with limited number of words.*

Keywords: K-means clustering, Malayalam, MFCC, Speech Recognition.

1. Introduction

Malayalam is a language spoken in India predominantly in the state of Kerala. It is one of the 22 scheduled languages of India with official language status in the state of Kerala and the union territories of Lakshadweep and Pondicherry. It belongs to the Dravidian family of languages, and was spoken approximately by 33 million people according to the 2001 census. Malayalam is also spoken in the Kanyakumari and Coimbatore districts of Tamil Nadu, southern parts and Kodagu districts of Karnataka State. It is also used by a large population of Indian expatriates living around the globe, including Persian Gulf, United states, Singapore, Australia and Europe. Malayalam language script consists of 53 letters with 37 consonants and 16 vowels. It is a syllable based language written with syllabic alphabet in which all consonants have an inherent vowel /a/ [1].

Speech or Voice Recognition is the computing task of validating a voice identity by using characteristics extracted from the voice. Speech is the most basic, common and efficient form of communication method for people to interact with each other. People are comfortable with speech therefore persons would also like to interact with computers via speech, rather than using primitive interfaces such as keyboards and pointing devices. This can be accomplished by developing an Automatic Speech Recognition (ASR) system, which allows a computer to identify the words that a person speaks into a microphone or telephone and convert that into written text [3].

In a multilingual society like India, where there are about 1672 dialect of spoken forms, Automatic Speech Recognition (ASR) technology has a wider scope. It would be a vital step in bridging the digital divide between English speaking Indian masses and others. Since there are no input

standards for Indian languages, it eliminates the keyboard mapping of different fonts of Indian languages [3].In the current Indian context; the machine-oriented interfaces restrict the computer usage to miniature fraction of the people, who are both computer literate and conversant with written English. This paper formulates a basic method to develop a basic speech recognition system for Malayalam, having the ability to recognize the isolated words in Malayalam speaker independently. Even though in this paper we are experimenting with different Malayalam words, this method can be used for other languages also. This method can be used for various speech based applications especially based on limited number of voice commands.

2. Literature Survey

Malayalam has 52 consonant phonemes, encompassing 9 places of articulation and 8 manners of articulation. In terms of manner of articulation, plosives are the most complicated, for they demonstrate a five-way distinction in velars, palatals, retro flexes, dentals and labials[12] [15].

Unlike English, Alphabets of Malayalam are classified according to the mechanism of speech production through our vocal system [13]. The vowels are realized as letters which origins from the vocal cord. Consonants are realized as sounds formed by the action of vocal tract. The consonants are divided into five groups each containing five letters. These five sections are grouped in accordance with which part of vocal tract is performing the production of that group of sounds. And more than that, each letters in these five groups differ with the consequent letter in a similar fashion. Thus the place of a particular letter in the whole list

of Malayalam alphabets itself says about the parts (or mechanism) of our vocal system involved in production of that sound. In this experiment we develop a basic speech recognition ability to identify the isolated spoken words in Malayalam [16].

In spite of the urgent need for automation in all domains, the development of strategies for speech recognition in regional languages is still perceived to be cumbersome due to various issues such as non-availability of speech corpus for training purpose, complexity in the language, lack of phoneme recognizers and difficulty in creating a speech corpus with necessary transcriptions. Though this ordeal is in focus over a period of time, it still remains a challenge and efforts are required to accomplish this with greater precision and reliability [6]. In a multilingual society like India only little work was done in making speech recognition systems in native languages. Major work is done on Hindi [4], Marathi [5] and Tamil [6] languages.

A strategy for recognizing spoken words in Malayalam language is an ongoing task. Currently several feature extraction techniques based systems have been proposed and a several experiments with these systems is going on. Speech recognition of isolated Malayalam words using wavelet features and artificial neural network [9] and Feature extraction in the time-frequency domain is performed using Wavelet Packet Decomposition (WPD) for Malayalam word recognition [2] are some of the work related to this area. A scheme is also proposed which uses a standard implementation, with some modifications to the noise detection/elimination algorithm and the HMM training algorithm [11].

3. Methodology

The various types of feature extraction and speech recognition approaches used here are explained in the following section.

3.1 Speech Recognition

In speech recognition where the mission is to recognize the speech (what is being said.). The verification or identification task is a 1:1 matching that consists of making a decision whether a given voice sample is produced matches the stored one. An identity claim (e.g., a PIN code) is given to the system, and the unknown voice sample is compared with the claimed stored voice template. If the degree of similarity between the voice sample and the template exceeds a predefined threshold, it is accepted, and rejected otherwise [14].

3.2 Feature Extraction

Feature extraction is a special form of dimensionality reduction, we will do that by extracting a specific features from the speech, these features carry the characteristics of the speech which are different from one speaker to another, so these features will play the major role as our mission is to identify the spoken word and make a decision that highly depends on how much we were successful in extracting a useful information from the speech in a way enables our

system to differentiate between different words and identify them according to their features.

3.3 MFCC

Several feature extraction algorithms are used, such as; linear predictive coefficients (LPC), linear predictive cepstral coefficients (LPCC), mel frequency cepstral coefficients (MFCC), and human factor cepstral coefficient (HFCC). We used the (MFCC) algorithm to extract the features, we choose it since it shows high accuracy results for clean speech [8]. MFCCs are less susceptible to the physical conditions of the speaker's vocal tract. The feature extraction is usually a non-invertible (lossy) transformation, as the MFCC described pictorially in Figure 1.
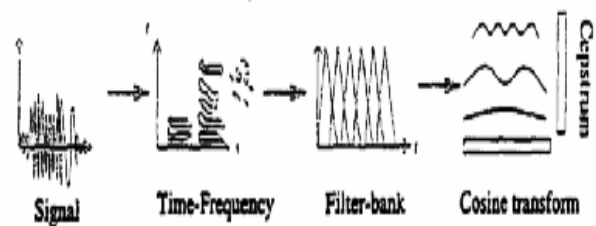

Figure 1: Pictorial representation of mel-frequency cepstrum (MFCC) calculation.

Figure 2 shows the block diagram for the feature extraction processes applying MFCC algorithm [7].
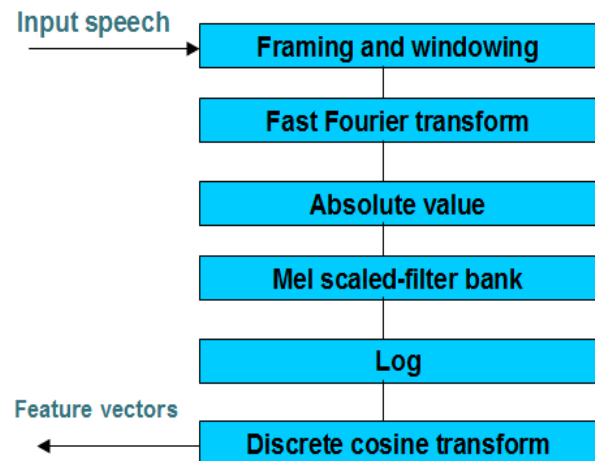

Figure 2: MFCC Flow diagram

3.4 Vector Quantization

Vector Quantization is a quantization technique used to compress the information and manipulate the data such in a way to maintain the most prominent characteristics [11]. VQ is used in many applications such as data compression (i.e. image and voice compression), voice recognition, etc. The technique of VQ consists of extracting a small number of representative feature vectors as an efficient means of characterizing the speech specific features. The clustering is done by a clustering algorithm.

3.5 K-means

The K-means algorithm partitions the T feature vectors into M centroids [17]. The algorithm first randomly chooses M cluster-centroids among the T feature vectors. Then each feature vector is assigned to the nearest centroid, and the new centroids are calculated for the new clusters. This procedure is continued until a stopping criterion is met, that is the mean square error between the feature vectors and the cluster centroids is below a certain threshold or there is no more change in the cluster-centre assignment. Flow diagram of the K-means algorithm is shown in figure 3[7].

3.6 Feature Matching

In the recognition phase an unknown voice, represented by a sequence of feature vectors; it is compared with the codebooks in the database. For each codebook a distortion measure is computed, and the one with the lowest distortion is chosen. This match is selected as the identified voice.
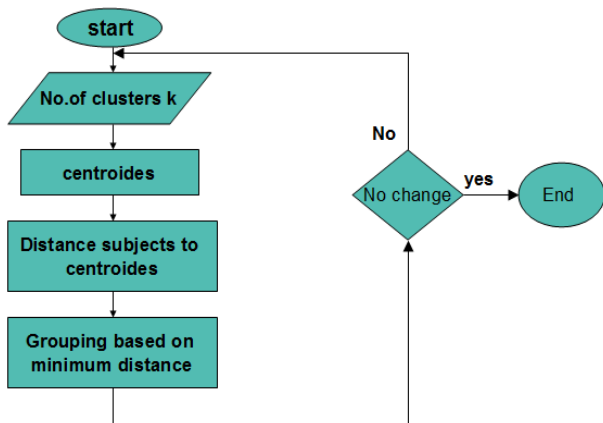
Figure 3: K means algorithm

4. Identification System

The system consists of two phases. In the training phase the sample words are stored in the database and in the testing phase different speakers are allowed to verify the voice commands one by one. The steps carried out during the testing phase can be summarized as figure 4.
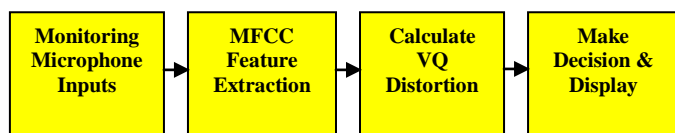
Figure 4: Identification phase

5. Speech Database

For conducting this experiment we choose six Malayalam words. The samples stored in the database are recorded by using a high quality studio-recording microphone at a sampling rate of 8 KHz.

SL .NO	Words	In English
1	ഒന്ന്	*Onnu*
2	രണ്ട്	randu
3	മൂന്ന്	*moonnu*
4	നാല്	*naalu*
5	അഞ്ച്	*anchu*
6	ആറ്	*aaru*

Table 1: Speech database

These six words are recorded and stored in a database in the training phase.
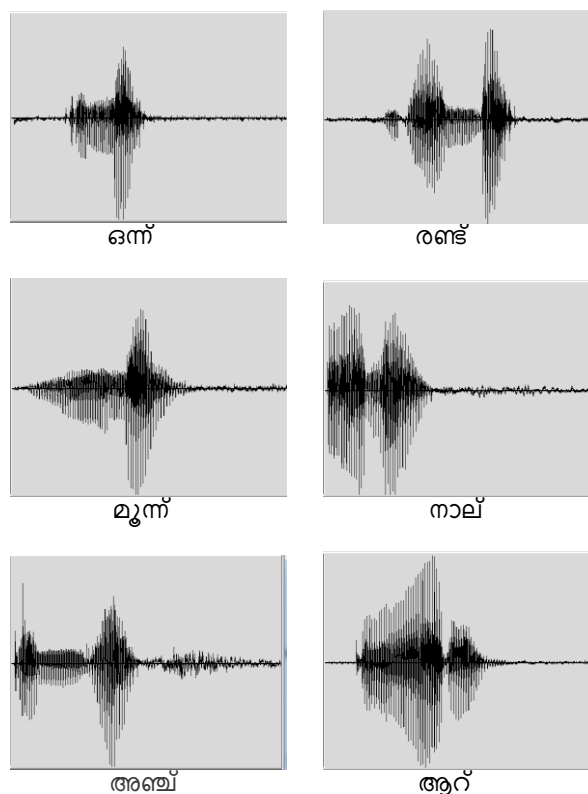
Figure 5: Wave forms of stored words

6. Training and Result

Table 2 shows the experimental results for ten different users. The speakers are asked to speak the stored words and the identification rate is noted down.

Speaker	ഒന്ന്	രണ്ട്	മൂന്ന്	നാല്	അഞ്ച്	ആറ്
1	C	C	C	C	C	C
2	X	X	C	C	C	C
3	C	C	C	C	C	X
4	C	C	C	C	C	C
5	C	C	C	C	C	C
6	C	C	C	C	X	C
7	X	C	X	C	C	C
8	C	C	C	C	C	C

9	C	C	X	C	C	C
10	C	C	C	C	C	X

Table 2: Sample identification phase result for ten speakers (C - identified, X - Not identified).

The accuracy obtained by this system is shown in Table 3. Further improvement can be obtained by a better VQ codebook design, with the training set including utterances from a large number of speakers with variation in ages, gender and accents.
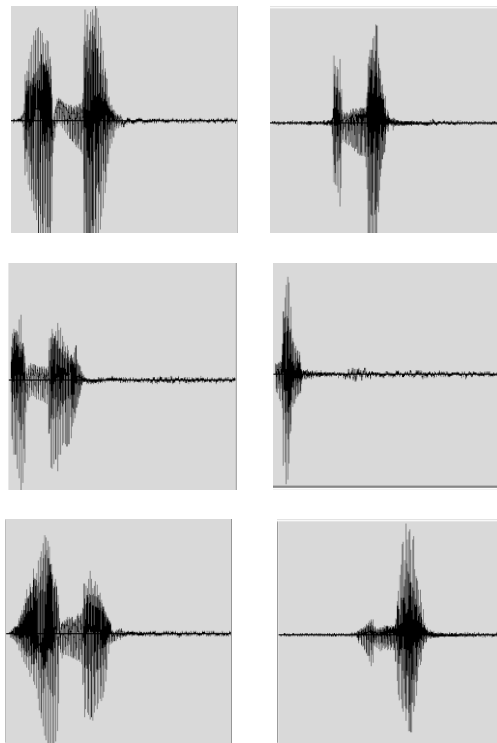
Figure 6: Different speaker's waveform of 'ഒന്ന്'

No.	Words	Accuracy
1	ഒന്ന്	85 %
2	രണ്ട്	93 %
3	മൂന്ന്	83 %
4	നാല്	92 %
5	അഞ്ച്	95 %
6	ആറ്	82 %
	Average	88 %

Table 3: Accuracy test results

The system consists of six stored words and hundred speakers and the experiment was conducted with almost 88 % accuracy.

7. Conclusion

Speech recognition remains one of the most important problems in the current scenario. It has numerous applications including those in, various sorting applications, mobile telephony, automatic scoring of tests containing multiple choice questions, automatic machine control, job application form sorting and various voice applications. The research in the domain of audio and speech recognition in various languages is still going on and is a challenging task .
To the best of our knowledge, little work has been done in Indian language, especially in Malayalam as compared with those for non Indian languages. This paper has discussed an effective method for recognition of isolated Malayalam words. This method can be also used for limited number of voice recognition in different languages irrespective of speakers. This confirms that the goal of creating a large vocabulary continuous transcription system is a realizable task in the near future.

References

[1] Cini Kurian, Kannan Balakrishnan *Automated Transcription System for Malayalam Language*, International Journal of Computer Applications (0975 – 8887) Volume 19– No.5, April 2011

[2] Sonia Sunny , David Peter, *Development of Speech Recognition System for Speaker Independent Isolated Malayalam Words* , International Journal of Computer Science & Engineering Technology (IJCSET), ISSN: 2229-3345 Vol. 3 No. 4 April 2012

[3] Vimala.C, Dr.V.Radha, *A Review on Speech Recognition Challenges and Approaches*, World of Computer Science and Information Technology Journal (WCSIT), ISSN: 2221-0741, Vol. 2, No. 1, 1-7, 2012

[4] Optimum HMM combined with vector quantization for hindi speech word recognition, Poonam Bansal, Amita Dev, Shail Bala Jain ,IETE journal of research Year : 2008 , Volume : 54 , Issue : 4 , Page : 239-243

[5] Siddheshwar S. Gangonda, Dr. Prachi Mukherji , *Speech Processing for Marathi Numeral Recognition using MFCC and DTW Features* , International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 National Conference on Emerging Trends in Engineering & Technology (VNCET-30 Mar'12)

[6] Spoken Word Recognition Strategy for Tamil Language AN. Sigappi1 and S. Palanivel, International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, January 2012,ISSN (Online): 1694-0814

[7] Allam Mousa ,MareText Independent Speaker Identification based on K-mean Algorithm, International Journal on Electrical Engineering and Informatics - Volume 3, Number 1, 2011

[8] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman ,*Speaker identification using Mel Frequency Ceptral Coefficients*, 3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh

[9] 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

[10] http://www.spectrogramsforspeech.com/tutorials-2/wavesurfer-overview/

[11] Kayte Charansing Nathoosing , *Isolated Word Recognition for Marathi Language using VQ and HMM*, Science Research Reporter 2(2):161-165, April 2012

[12] Tara Mohanan , *Syllable Structure in Malayalam*, , Linguistic Inquiry, Vol. 20, No. 4 (autumn, 1989), pp. 589-625

[13] http://www.omniglot.com/writing/malayalam.htm (Visited September 2012)

[14] Daniel Jurasfky and James H. Martin, *An introduction to natural language processing, computational linguistics, and speech recognition*, Pearson Education, Inc. 2000

[15] Haowen Jiang, *Malayalam: a Grammatical Sketch and a Text,* Department of Linguistics, Rice University

[16] http://code.google.com/p/sharika (visited September 2012)

[17] http://www.bio.umontreal.ca/casgrain/en/labo/k-means.html (visited September 2012)

[18] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html (visited September 2012)