International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 10, October 2017
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A Database of Marathi Numerals for Speech Data Mining

Ms. Himgauri Kondhalkar[1] , Mrs. Prachi Mukherji[2]

[1]Research scholar , E & Tc Department, Sinhgad College of Engineering, Pune, (India)

[2]Research Guide , E & Tc Department, Cummins College of Engineering, Pune, (India)

ABSTRACT

This paper presents a versatile dataset of numerals used in Marathi language. The corpus consists of Marathi numerals spoken by variety of speakers over a microphone. The dataset can be used for development of speech recognition system for Automated Teller Machine or lifts for visually impaired people across Maharashtra where Marathi is widely spoken language. There are various datasets available for languages like English, German but very few for Indian languages. The accuracy of entire speech recognition system is dependent upon the quality of data collected for training and testing purpose. Speech data of 100 participants was recorded in a closed environment. The dataset can be made available to encourage other researchers to use it for testing their own speech recognition methods.

Keywords- Marathi numerals, speech database, speech recognition

I.INTRODUCTION

Researchers have tried to develop system to make computer record, interpret and understand human speech. Marathi language is spoken mainly in central and western India. Such systems are useful in different areas like agriculture, security, health care, commerce etc. Speech technologies can play a very important role in development of applications for common people in a multi-lingual society such as India which has about 1652 dialects/native languages [1]. The dataset described in this work mainly focuses on developing a speech recognition system for visually impaired people. There are a number of Automated teller machines available for such people where European languages are used. This work proposes to build a human computer interaction system using Marathi, an Indo European language. The system developed is speaker independent. Since the dialect varies from person to person, varied samples need to be collected in terms of gender and dialect. The current dataset is recorded with the goal of creating an efficient speech recognition method for Marathi speech. This paper is organized as follows: The second section explains Marathi language related literature. The third section describes related work in the same domain. The fourth section explains the database collection. Next section describes the voiced part separation of the recorded sample. Finally the results of classification are discussed.

II. MARATHI LANGUAGE

Marathi is an Indo-Aryan language, spoken by the people of western and central India [2]. Marathi is derived from Devnagari script used for writing Sanskrit. Sanskrit is the origin for languages like Marathi, Hindi. Maharashtra state is second most populous state of India with 36 districts. Marathi is the official language of Maharashtra. Marathi has almost 42 dialects like standard dialect spoken by native people of Maharashtra, Thanjavur dialect spoken by people who migrated from Karnataka and Andhra Pradesh , Bhavsar Marathi, Vaidarbhi dialect etc. Marathi language has 12 vowels and 36 consonants. Each word in Marathi is formed with the combination of vowels and consonants. For example, the numeral “ek”(one) is pronounced as ‘/e k/’ which is a combination of a vowel ‘e’ and consonant ‘k’. Marathi words have a lot of pronunciation inconsistency as the

pronunciation unlike English are related to the context of the word. Generation of Marathi speech database considering all the above diversities is a crucial stage in a speech recognition system.

III.RELATED WORK

Speech databases have been prepared for text to speech or speech to speech conversion. There are some databases recorded in Marathi language for different applications. Following paragraph provides brief information about such databases. P. P. Shrishrimal et al. have collected speech data from 100 speakers. Each speaker will be asked to speak 100 words with 5 utterance of every word. Total 500 utterances of the words will be collected from every speaker with 100 words with 5 utterances each. The speech corpus was generated using various agriculture related websites [2]. TIFR (Mumbai) and IIT Bombay together executed a project for speech data collection over phone line. The project is sanctioned by the Technology Development for Indian Languages (TDIL) for the development of Speech Recognition system for agriculture purpose. The speech database will consist of data recorded from approximately 1500 speakers along with background noise [3]. A Large Vocabulary Marathi Continuous Speech Database is developed at IIIT Hyderabad with coordination of HP Labs Bangalore. Speakers are from different age groups and native speakers of the language. The dataset contains noisy speech samples [4]. Bharati Gawali, Santosh Gaikwad et al. developed a Marathi speech dataset with 105 samples of Marathi vowels, 420 samples of isolated words and 175 samples of sentences. The aim was to compare the performance using MFCC and DTW [9].

IV.SPEECH DATABASE COLLECTION

Following are the selection criteria used for generating the proposed database.

- Selection of age group- Speakers speaking Marathi with age groups between 18years to 25 years, 26 years to 40 years and 41 years to 60 years are selected for recording the database. Total 100 speakers with 40 female speakers and 60 male speakers were selected.
- Vocabulary size of the database- since automated teller machine is the target application for the speech data collection, Marathi numerals are recorded. Marathi language has 10 basic numerals as shown in table 1. Thus the vocabulary size of the database is Marathi numerals from “shunya” to “nau”.

Table 1:Marathi numerals recorded and their pronunciation

Numeral	०	१	२	३	४	५	६	७	८	९
pronounciation	‘Shunya’	‘ek’	‘don’	‘tin’	‘chaar’	‘paach’	‘sahaa’	‘saat’	‘aath’	‘nau’

- Selection of dialect- out of the 100 speakers selected , 60 speakers are native people of Maharashtra with fluency in speaking standard dialect Marathi. Out of remaining 40 speakers, 25 speakers have different dialect like Varhadi, Tanjavur, and Konkani. Remaining 15 from this set were speakers who were not from Maharashtra speaking languages like Punjabi, Tamil etc. These speakers have migrated from other states to Maharashtra for some years. They were asked to utter Marathi numerals.

V.RECORDING SPEECH DATA

The speech data was recorded in a closed room with proper arrangements to avoid reverberation of sound. The recording is done using professional SM 58 LC SHURE microphone. Distance between the speaker’s mouth and microphone is maintained to 10cm. The recording time for each utterance was approximately 1 second. Speaker was instructed to utter each numeral 10 times in a sequence. The recorded samples are stored in .wav format. Each wav file has 100 utterances with 10 utterances of each numeral. Each utterance had a separate instance of recording. Thus total size of the database is 10000 samples. The recording was done with Audacity software. It is free, open

source software available with latest version of 2.1.2 which can run on windows operating system platform and designed for recording and editing sounds. Sampling rate used for recording was 16Khz. To segment the recorded samples in a .wav file and separate voiced part from unvoiced part of each speech sample, short term energy is calculated. Fig.1 shows the recorded speech sample with duration of one second. It shows the voiced and unvoiced part both. The recording is clean with minimum background noise. Each recorded numeral is segmented into number of frames with 50% overlap. Fig. 2 shows the voiced part of the speech sample separated using short term energy threshold.

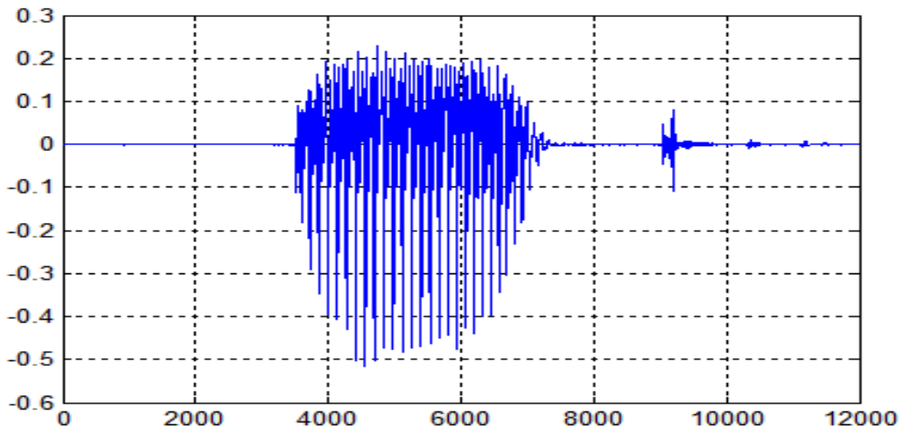

Fig. 1 Recorded speech for numeral "ek"

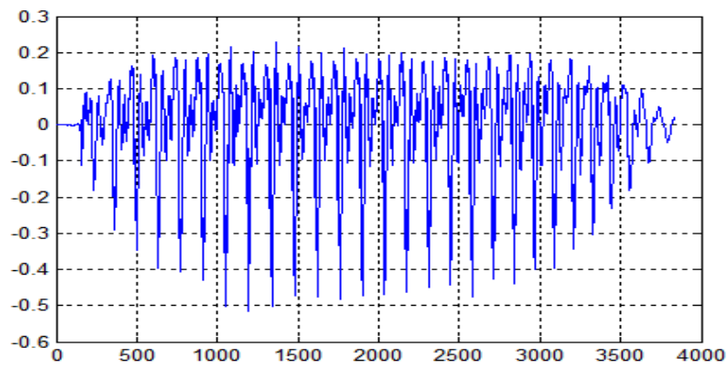

Fig. 2 Voiced part for numeral "ek"

Framing of the recorded speech is done with frame length calculated as in equation (1).

$$s_L = \frac{length(y)}{8} \tag{1}$$

Where s_L is the frame length that is variable depending on length of the input speech sample. A 50% overlap is considered between two consecutive frames as in equation (2). Then the total number of frames F is calculated using equation (3).

$$s_1 = \frac{s_L}{2} \tag{2}$$

$$F = \frac{length(y) - w_L}{s_1} + 1 \tag{3}$$

Short term energy (STE) is calculated per frame using equation (4).

$$e(n) = \sum_{m=1}^{N} (s(m).w(n-m))^2 \tag{4}$$

Where $e(n)$ is the short term energy, N is total number of samples per frame. Deciding the threshold value for separating voiced and unvoiced part from the recorded speech is a critical part of the preprocessing. In the proposed work threshold is calculated using (5).

$$T = \frac{\sqrt{\sum_{i=1}^{N} s(i)^2}}{s_L} \quad (5)$$

Where T is the threshold value. $e(n) \geq T$ is considered as the voiced part of the recorded speech and values of STE below threshold are ignored. Thus voiced part of the speech sample is obtained which is further used in speech recognition. Fig. 3 shows the voiced part of the speech sample divided into number of frames.

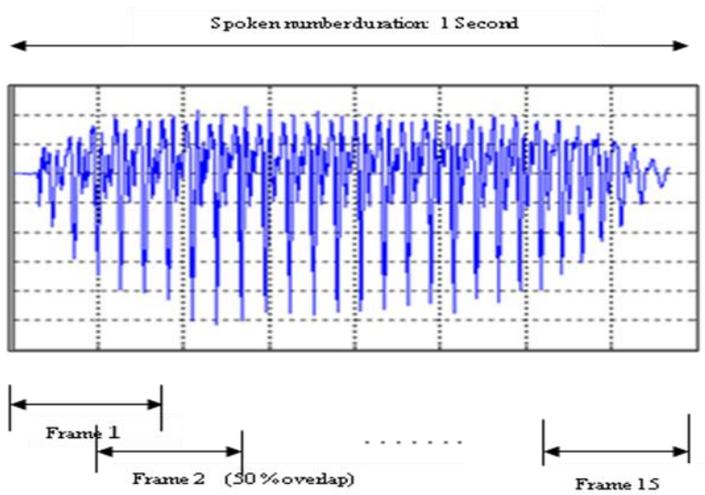

Fig. 3 Speech sample after framing for the numeral 'ek'

All the speech samples are manually checked to ensure that each of the recording matches with the expected numeral with minimum background disturbance.

VI. RESULTS

The dataset thus prepared is further used for feature extraction and classification. Table 2 lists the features extracted from recorded data. These are commonly used in speech processing.

Table 2: List of extracted features

Category	Features Extracted
MFCC	Mel frequency cepstral coefficients(210 features)
BFCC	Bark scale and DCT coefficients(210 features)
Energy	Average energy per frame
Zero crossing rate	Average zero crossing rate per frame

Classification results using Artificial Neural Network are shown in fig. 3. The results prove that dataset prepared is versatile as well as accurate.

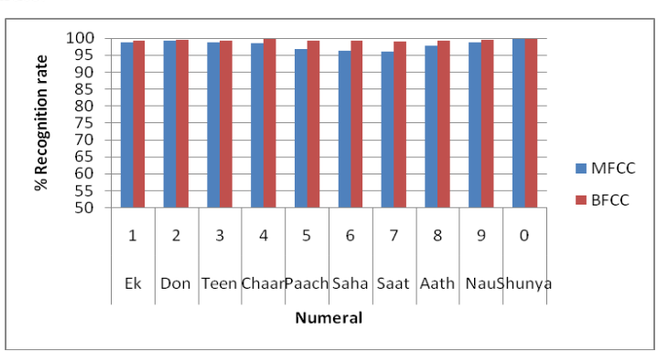

Fig. 3 Recognition rate in % for Marathi numerals.

VII.CONCLUSION

This paper presents Marathi language speech corpus. The size of the corpus is 10000 speech samples recorded for ten Marathi numerals. The dataset covers variety in terms of gender, age and dialect spread across Maharashtra. The speech corpus can be used for applications like Automated teller machines, lifts for visually sighted people. It can be used for hands free dialing systems for phones, search engines etc. Hence, the corresponding speech data should be valuable for training speaker independent, isolated speech recognition systems for Marathi language. The dataset can be made available for other researchers to try their speech recognition algorithms.

REFERENCES

[1] Ratnadeep Deshmukh, Dr. Vishal Waghmare, Indian Language Speech Database: A Review, *International Journal of Computer applications, 47(5),* June 2012, 17-21.

[2] P.P. Shrishrimal, Ratnadeep R. Deshmukh, Dr. Vishal Waghmare, Development of Isolated Words Speech Database Of Marathi Words For Agriculture Purpose, *Asian Journal of Computer Science and Information Technology, 2(7),* 2012, 217-218.

[3] Tejas Godambe, Namarata Karkera, Samudravijaya K., Adaption of Acoustic models for Improved Marathi Speech Recognition, Acoustics 2013, 1-6, November 2013.

[4] Kishore Prahallad, E. Naresh Kumar, et al., The IIIT-H Indic Speech Databases, Interspeech, Portland, USA, 2012

[5] Yogesh Gedam, Sujata Magare, Amrapali Dabhade, Ratnadeep Deshmukh,Development of Automatic Speech Recognition of Marathi Numerals, *International Journal of Engineering and Innovative Technology, 3(9),* March 2014.

[6] Mansi Kolwankar, Manasi Dixit, "English to Marathi Audio Dictionary using Speech Recognition", fifth IRAJ International Conference, pp. 57-59, Pune, September 2013.

[7] Tejas Godambe and Samudravijaya K., Speech Data Acquistion for Voice based Agricultural Information Retrieval , 39th All India DLA Conference, Punjabi University, Patiala, June 2011.

[8] Potale Shubham, Kharpude Pratik, Patil Rahul, Ajay Kumar Gupta, Design and Development of Word Recgnition for Marathi Language, *Imperial Journal of Interdisciplinary Research, 3(6),* 2017,30-32.

[9] Bharati Gawali, Santosh Gaikwad, Pravin Yannawar, Suresh Mehrotra, Marathi Isolated Word Recognition System using MFCC and DTW Features, *International Journal on Information Technology, 1(1),* 2011, 21-24.

[10] P.P. Shrishrimal, Ratnadeep R. Deshmukh, Ganesh Janwale, Devayani Kulkarni, Marathi Digit Recognition System based on MFCC and LPC Features, *International Journal of Advanced Innovative Research,6(6),* 2017, 37-39.