Marathi Digit Recognition System based on MFCC and LPC features

Pukhraj P. Shrishrimal[#1], Ratnadeep R. Deshmukh[#2], Ganesh B. Janwale[*3], Devyani S. Kulkarni[*4]

[#] *Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad - 431004, (MS), India*

[1]*pukhraj.shrishrimal@gmail.com*

[3]*rrdeshmukh.csit@bamu.ac.in*

[3*]*MGM's Institute of Biosciences and Technology, Aurangabad -413 003 (MS) India*

[2]*ganesh.janvale@gmail.com*

[4*], *Xpanxion International Private Limited, Sarjaa Rd, Sahil Park, Sanewadi, Aundh, Pune-411007, (MS), India*

[4]*devyanikulk @gmail.com*

***Abstract*— This paper presents the Marathi digit recognition system based on MFCC and LPC feature using confusion matrix. The speech database used for this work is unique as it consists the speech samples recorded in regular environment with background noise and with a huge variation in the pronunciation as the speakers belongs to different places around the Aurangabad city. The recognition rate achieved is less but it is a first attempt for developing the digit recognition system using the developed speech database.**

***Keywords*— Automatic Speech Recognition, Speech Database, Mel Frequency Cepstral Coefficient, Linear Predictive Coding.**

I. INTRODUCTION

The development of language technologies is playing an important role in the digital society. Human beings are trying to develop different interface system to communicate with computer system. Speech being the most widely used mode of communication between human; so researchers are motivated to develop speech based system [1]. Researchers around the world are working in the domain of speech recognition for long time and the result of which can be seen as implementation in many devices.

However the development of these devices is basically in specific languages and they can be used in those languages only. The work for Indian languages in terms of application orientation is less. The reason for this lack is different languages and dialects spoken in same state. The speech based system can play a vital in the development of multilingual society like India. The language technology can help to bridge the gap between technically illiterate people to join the main stream and be part of Digital India and avail the benefits of it.

In this paper we have presented the Speech recognition system developed for the numbers zero to nine (0 to 9) in Marathi language using MFCC and LPC. This work is an attempt to develop speech based system for which the data is recorded in normal environment with background noise. This paper is organized as follows: section II gives a brief idea about Marathi language; section III gives the details of the database used for the experiment; section IV describes what is meant by feature extraction; Section V explain the basics Mel Frequency Cepstral coefficient (MFCC) and Linear Predictive Coding; Section VI presents the results for recognition of spoken digits; section VII outlines the conclusion and future work.

II. MARATHI LANGUAGE

Marathi is one of the 22 officially recognized language by the constitution of India. There are 74,775,760 users of Marathi language around the world. Marathi is the official language of Maharashtra and co-official language in the union territories of Daman and Diu and Dadra and Nagar Haveli. It is spoken in Maharashtra state and at few other places in states like Andra Pradesh, Goa, Karnataka, Madhya Pradesh and Chhattisgarh [2]. Indic scholars have recognized 42 different dialects of Marathi language. Marathi language was initially written using the Modi script; since 1950 it is written in Devanagari script [3].

III. SPEECH DATABASE

The Speech Database used for the research consist of 5000 utterances of the numbers from 0 to 9. The speech samples are collected from 100 native Marathi speakers. Each speaker was asked to speak each number five times. From the 100 speakers 50 speakers were male and 50 were female. The speech data was captured in a regular room with a sampling frequency of 16,000 Hz in 16 bit mono and stored in .wav format. The speaker belong to different villages around the city of the Aurangabad.

IV. FEATURE EXTRACTION

Feature extraction is a basic and fundamental pre processing step to pattern recognition and machine learning problem. It is a special form of dimensionality reduction technique used to reduce the data which is very large to be processed by an algorithm. In feature extraction the provided input data is transformed into a set of features which provides the relevant information for performing a desired task without the need of the full size data but using the reduced set.

The speech recognition technique is having a background from the DSP i.e. Digital signal processing. DSP has been the centre of progress in speech processing during the complete development of the speech processing and speech recognition systems [4]. It is not only used in speech analysis, synthesis, coding, recognition and enhancement but also in voice modification, speaker recognition and language identification.

Theoretically it is possible to recognize speech directly from the digital waveform of the speech. However, as speech is time varying the idea to perform some form of feature extraction came into existence which is used to reduce the variability of speech signal. In the context of Automatic speech recognition feature extraction is the process of retaining the useful information from the speech signal while the unnecessary and unwanted information is removed which involves the speech signal analysis. However, while removing the unwanted information from the speech signal some useful information may also lose [5].

The objective to be achieved with feature extraction is to untangle the speech signal into the different acoustically identifiable components and to obtain the set of feature with low rate of change in order to keep the computation feasible. The feature extraction for speech recognition can be divided in spectral analysis, parametric transformation and statistical modelling.

A. Mel Frequency Cepstral Coefficient (MFCC):

The Mel Frequency Cepstral Coefficient is the well known and most widely used feature extraction method in speech domain. The MFCC is based on the human auditory perception system. The human auditory perception system does not follow a linear scale of frequency. For each tone with actual frequency f measured in Hz, a subjective pitch is calculated known as 'Mel Scale'. The mel frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1000Hz. As a reference point, the pitch of a 1 KHz tone, 40 dB above the perceptual hearing threshold is defined as 1000 Mels [6].

B. Linear Predictive Coding:

A linear prediction (LP) model [7] predicts/forecasts the future values of a signal from a linear combination of its past values. A linear predictor model is an all-pole filter that models the resonance (poles) of the spectral envelope of a signal or a system. LP models are used in diverse areas of applications, such as data forecasting, speech coding, video coding, speech recognition, model based spectral analysis, model-based signal interpolation, signal restoration, noise reduction, impulse detection, and change detection. In the statistical literature, linear prediction models are often referred to as autoregressive (AR) processes. The all-pole LP model shapes the spectrum of the input signal by transforming an uncorrelated excitation signal to correlated output signal whereas the inverse LP predictor transforms a correlated signal back to an uncorrelated flat-spectrum signal.

Inverse LP filter is an all-zero filter, with the zero situated at the same position in pole-zero plot as the poles of the all-pole filter and is also known as a spectral whitening, or de-correlation filter. Poles are at denominator of the polynomial and zeros are at numerator of the polynomial.

The all-pole LP model shapes the spectrum of the input signal by transforming an uncorrelated excitation signal to correlated output signal whereas the inverse LP predictor transforms a correlated signal back to an uncorrelated flat-spectrum signal. Inverse LP filter is an all-zero filter, with the zero situated at the same position in pole-zero plot as the poles of the all-pole filter and is also known as a spectral whitening, or de-correlation filter . Poles are at denominator of the polynomial and zeros are at numerator of the polynomial.

V. Experimental Analysis

The MFCC features of the database was calculated and the recognition was performed using confusion matrix. The recognition rate of the numbers for MFCC features using confusion matrix is shown in table 1. From table we can view that the lowest recognition for the number is 13.25% for the number eight. The highest recognition rate is achieved for the number zero which is 78.94%.

The LPC features of the database was calculated and the recognition was performed using confusion matrix. The recognition rate of the numbers for LPC features using confusion matrix is shown in table 2. From table we can view that the lowest recognition for the number is 12.32% for the number Seven. The highest recognition rate is achieved for the number one which is 66.17%.

TABLE I
Confusion Matrix of MFCC features of Spoken numerals zero to Nine

	0	**9**	**1**	**7**	**4**	**3**	**6**	**2**	**8**	**5**	**Total No. of Sample Tested**	**Recognition Rate in %**
0	**60**	3	8	0	1	0	0	1	1	2	76	78.94
9	1	**56**	0	0	2	0	2	14	1	1	77	72.72
1	9	0	**47**	1	0	10	0	2	0	0	69	68.11
7	0	0	0	**29**	12	0	14	2	15	9	81	35.80
4	6	9	2	6	**43**	0	5	1	4	0	76	56.57
3	3	1	11	0	0	**57**	0	2	0	0	74	77.02
6	0	3	0	14	4	0	**34**	6	5	9	75	45.33
2	1	23	0	0	1	0	1	**46**	0	0	72	63.88
8	0	2	0	38	12	0	12	1	**11**	7	83	13.25
5	0	5	0	13	6	0	20	1	10	**12**	67	17.91

The reason for such low recognition rate is the quality of speech database. The speech database used for the experiment was recorded in the regular environment and not in a noiseless closed environment. The other reason is the time alignment of the speech signals. Each speech sample is of different duration and when time alignment is performed the spectral features are varied due to either compression or decompression of the

speech signal. Even the background noise recorded in the all speech samples is different and the effect of a specific speech enhancement technique is not suitable.

TABLE III
CONFUSION MATRIX OF LPC FEATURES OF SPOKEN NUMERALS ZERO TO NINE

	4	**2**	**1**	**7**	**5**	**9**	**0**	**3**	**6**	**8**	**Total No. of Sample for Testing**	**Recognition rate in %**
4	**27**	2	0	9	13	6	3	1	4	19	84	32.14
2	0	**41**	5	0	0	20	0	0	0	7	73	56.16
1	0	5	**45**	0	0	4	6	5	0	3	68	66.17
7	17	3	0	**9**	15	3	1	0	16	9	73	12.32
5	6	1	1	6	**30**	5	0	0	11	8	68	44.11
9	0	19	3	1	0	**46**	2	0	0	4	75	34.5
0	4	1	15	0	2	1	**32**	22	3	4	84	26.88
3	0	2	15	0	0	0	13	**44**	0	3	77	33.88
6	12	1	0	6	9	5	1	0	**26**	9	69	17.94
8	11	1	0	5	6	15	3	0	0	**38**	79	30.02

Devyani et.al. developed a Marathi Isolated Speech Recognition System using HTK. They have used part of the same speech database with total 800 utterances of 40 individuals. They achieved a recognition rate of 99.75% with 48.75% accuracy at word level [8]. The accuracy of the developed system was low but the recognition rate was high due to language and acoustic models developed for the speech samples. This comparison proves that the speech recognition system developed using the training samples consisting different type of background noise and phonetic variation results in low recognition and accuracy of the system.

VI. CONCLUSIONS

This paper is an attempt to develop a speech recognition system based on the MFCC and LPC features for a speech database which is recorded in the normal noisy environment and which is having a huge variation as the speaker belongs to different location and the difference that is observed in the pronunciation of the speaker belonging to different places also affects the recognition rate. This paper has provided us a new objective how to develop robust speech recognition which will not be affected by the different noise which is observed and heard in the surrounding and how to tackle the difference for the phonetic language like Marathi where we can observe a slight change after every few kilometers.

ACKNOWLEDGMENT

This work is supported by University Grants Commission under the scheme Major Research Project entitled as "Development of Database and Automatic Recognition System for Continuous Marathi Spoken Language for agriculture purpose in Marathwada Region". The authors would also like to thank the Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad for providing the infrastructure to carry out the research.

REFERENCES

[1] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, Vishal B. Waghmare, "Indian Language Speech Database: A Review", International Journal of Computer Applications (0975 – 888) Volume 47– No.5, pp. 17-21, June 2012.

[2] https://www.ethnologue.com/language/mar cited on 17/04/2017 at 01:14 am

[3] http://www.omniglot.com/writing/marathi.htm cited on 17/04/2017 at 01:19 am

[4] Lawrence R. Rabiner and Ronald W. Schafer, "Digital Processing of Speech Signals, Signal Processing", Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[5] Bhupinder singh, Rupinder Kaur, Nidhi Devgun, Ramandeep Kaur, "The process of feature extraction in automatic speech recognition system for computer Machine interaction with humans: A Review", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 2, Feb 2012.

[6] Vibha Tiwari, "MFCC and its application in speaker recognition", International Journal on Emerging Technologies, Vol. 1, No. 1, pp. 19-22 (2010).

[7] R. Zelinski and P. Noll, "Adaptive transform coding of speech signals," IEEE Trans. Acoust. Speech Signal Process. Vol. ASSP-25, no. 4, pp. 299–309, Aug. 1977.

[8] Devyani S. Kulkarni, Ratnadeep R. Deshmukh, Vandana L. Jadhav Patil, Swapnil D. Waghmare, Pukhraj P. Shrishrimal, Aaron M. Oirere, "Marathi Isolated Digit Recognition System Using HTK", 2nd International Conference on Cognitive Knowledge Engineering 2016, Organized by Department of Computer Science and Information Technology, Dr. B. A. Marathwada University, Aurangabad, December 2016.

