IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. IV. (Jul.-Aug. 2017), PP 14-18
www.iosrjournals.org

LPC and HMM Performance Analysis for Speech Recognition Systemfor Urdu Digits

Shaikh Naziya S.[1], R. R. Deshmukh[2]
[1,2]*Department of CSIT, Dr. B.A.M. University, Aurangabad (MS), India*
Corresponding Author: Shaikh Naziya S.
naziyamsc@gmail.com

Abstract: *The research work concentrated on the development of digit corpus for the Urdu language, which starts from siffar (Zero) se Nau (Nine) and three utterances were taken in the development of corpus from 50 persons which were the native language speaker of Urdu. In that 50 speaker 50% male and 50 % female speakers were considered and all were having age above 18 years. Twoutterances were considered for training and oneutterance considered for testing.The performance analysis were done using the traditional techniques LPC (Linear Predictive Coding) and HMM (Hidden Markov Model). From the result analysis it has been conclude that Average System performance accuracy (PA) for Numeric data Zero to Nine (0-9) is 74% and WER is 26%. Number six and nine gives 100% result.*

Keywords: *Speech Recognition System, Linear Predictive Coding, Hidden Markov Model*

Date of Submission: 13-07-2017 Date of acceptance: 28-07-2017

I. Introduction

The aim of Automatic Speech Recognition research is to develop computational techniques to convert acoustic speech signals into strings of words. The problem of general speech recognition has not been solved yet for any language. This means that there exists no such versatile speech recognition system which can recognize the words spoken by any person, of any gender, in any accent and at any rate of speech in any environment. Therefore the task of speech recognition starts by making simplifying assumptions which manifest as the constraints on the system. In other words, all the open ended variables are constrained to certain practical limits which render the problem more feasible to tackle.

There are many areas where Automatic Speech Recognition (ASR) systems can play a pivotal role in facilitating the daily activities, and where the current levels of accuracy that these systems have attained can prove useful. One of these areas is speech based human computer interaction (HCI)[1]. This line of research promises to be of significant advantages in areas where keyboards may not be appropriate and natural language communication is desired. This includes control applications where hands and eyes may be busy at the same time and speech becomes a good means of issuing the commands. In addition to this, such systems can be of immense use for people with vision related disabilities, lack of motor control, crippled hands etc. In the under-developed countries where literacy rate is poor, this can provide a mechanism of information access to people who are unable to read and write as well as people who may be literate but not qualified in computing skills. Speech based HCI ideally brings computers within reach of anyone who can speak and listen.

II. About Urdu Language

Urdu (/ˈʊərduː/; Urdu: اُردُو ALA-LC: Urdū; IPA: [ˈʊrɖuː] or Modern Standard Urdu is a standardisedregister of the Hindustani language. It is the national language and lingua franca of Pakistan, and an official language of six states of India. It is also one of the 22 official languages recognized in the Constitution of India. Urdu is historically associated with the Muslims of the region of Hindustan. Apart from specialized vocabulary, Urdu is mutually intelligible with Standard Hindi, which is associated with the Hindu community. The Urdu language received recognition and patronage under the British Raj when the British replaced the Persian and local official languages of North Indian Jammu and Kashmir state with the Urdu and English language in 1837.

There have been attempts to "purify" Urdu and Hindi, by purging Urdu of Sanskrit loan words, and Hindi of Persian loan words, and new vocabulary draws primarily from Persian and Arabic for Urdu and from Sanskrit for Hindi. This has primarily affected academic and literary vocabulary, and both national standards remain heavily influenced by both Persian and Sanskrit. English has exerted a heavy influence on both as a co-official language [2].

III. Literature Review

3.1 Automatic Speech Recognition System (ASR System)

Speech recognition (SR) is the translation of spoken words into text. It is also known as "automatic speech recognition or ASR", "computer speech recognition", "speech to text or STT". Speech Recognition is an inter-disciplinary research domain. Speech Recognition (is also known as Automatic Speech Recognition (ASR) or computer speech recognition) is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program [3]. Figure 1 shows Basic Model of speech Recognition.

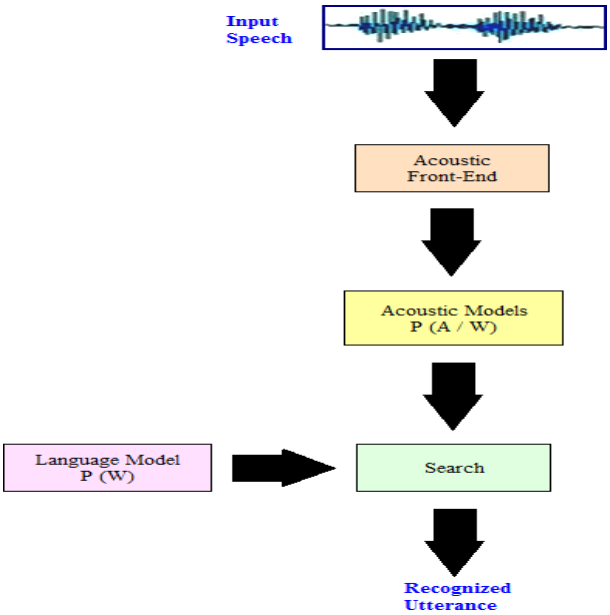

Figure 1: Basic Model of Speech Recognition

3.2 Linear Predictive Coding (LPC)

One of the most powerful signal analysis techniques is the method of linear prediction. LPC [4–5] of speech has become the predominant technique for estimating the basic parameters of speech. It provides both an accurate estimate of the speech parameters and it is also an efficient computational model of speech. The basic idea behind LPC is that a speech sample can be approximated as a linear combination of past speech samples. Through minimizing the sum of squared differences (over a finite interval) between the actual speech samples and predicted values, a unique set of parameters or predictor coefficients can be determined. These coefficients form the basis for LPC of speech [6]. The predictor coefficients are therefore transformed to a more robust set of parameters known as Cepstral coefficients. Figure 2 shows the steps involved in LPC feature extraction.

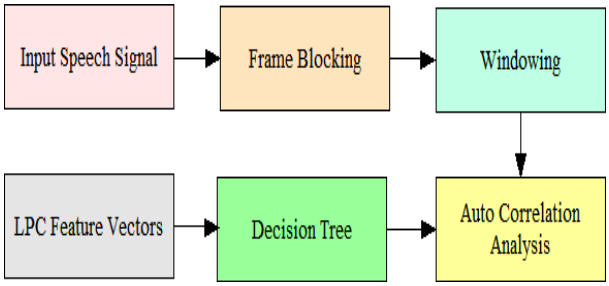

Figure 2: Steps involved in LPC Feature Extraction

3.3 Hidden Markov Model

It is widely use stochastic approach today because in this approach speech is generating from number of states for each HMM model. In HMM we mixture multi vibrate Gaussian distribution, probabilistic mean, variance and mixture weight for speech [7]. Each phoneme has different output distribution. A HMM for a sequence of words or phoneme is made by concentrating the individual train HMM for the separate words and phoneme. Figure 3 shows the information about the hidden markov model (HMM)

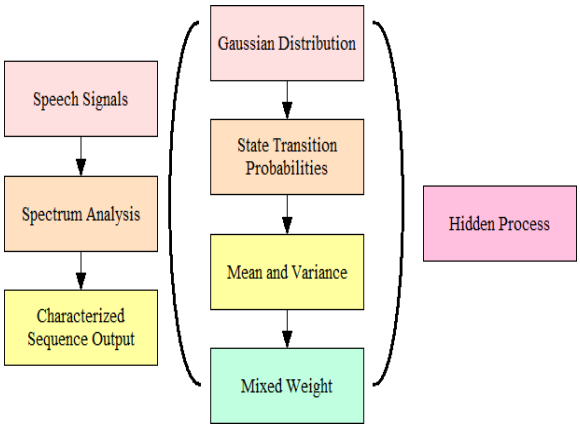

Figure 3: Hidden Markov Model (HMM) Block Diagram

1. Development Of Database

The Database was developed numeric speech dataset which starts from siffer to nau (Zero to Nine) total 10 digits ofthe numerous. There are total 50 speakers 35 female and 15 male. Three utterances are taken of each word. Total utterances are 1500. Recording frequency is 16000Hz. File is saving in mono typeand .wav format.

Table 1: Numeric Speech Corpus

Sr. No.	Urdu Words	Pronunciations		Meaning in English
1	(۰) سـ فر	सिफर	Sifar	Zero
2	(۱) ایـ ک	ऐक	Aik	One
3	(۲) دو	दो	Do	Two
4	(۳) تـ ین	तीन	Teen	Three
5	(٤) چار	चार	Chaar	Four
6	(٥) پـ انـ چ	पांच	Paanch	Five
7	(٦) چھ	छ	Chha	Six
8	(۷) سات	सात	Saat	Seven
9	(۸) آٹھ	आंठ	Aanth	Eight
10	(۹) نـ و	नौ	Nau	Nine

3.4 Feature Extraction And Classification

We use LPC and HMM techniques for analysis of database, feature extraction and recognition. Before the techniques were selected we studied about various techniques and from the literature review we have choose the techniques which are suitable for our system due to their better performance. Analysis were done using features LPC and classification by HMM technique. Figure 3 shows the proposed basic diagram of Urdu speech recognition from acquisition of speech to testing.

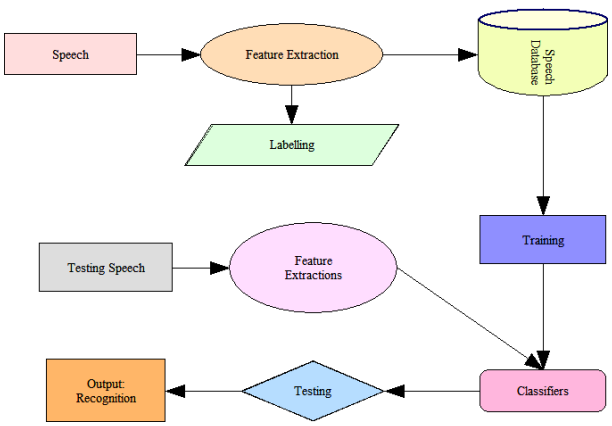

Figure 3: Block diagram of Speech Recognition System for Urdu

IV. Performance Analysis

This section deals with the performance of the whole system from the feature extracted to the classification and the recognition. The features used in our system are Linear Prediction Coding (LPC) and for classification Hidden Markov Model (HMM) is used. We also discuss here Performance accuracy of the system and Word Error Rate (WER).

4.1 Evaluation Criteria

The evaluation was made according to recognition accuracy and computed using Word Error Rate (WER).

4.1.1 Performance Accuracy (PA)

Individual accuracy is calculated as shown in equation 1:

$$Performance\ Accuracy = \frac{S}{N} * 100 \tag{1}$$

Where, S = Number of Recognized Sample,

N = Total Number of Sample

4.2.2 Word Error Rate

WER defined by the equation 2 below, which aligns a recognized word string against the correct word string and compute the number of substitution (S), deletion (D), and insertion (I) from the number of correct words (N):

$$WER = \frac{(S + D + I)}{N} * 100 \tag{2}$$

WER of 20% means that there was an error (substitution, deletion or Insertion) of 20 words out of 100 words spoken by a speaker.

HMM applied on LPC of male and female speakers on numeric speech data (0 to 9)

LPC features have been extracted for every uploaded wave file and then database of these features is prepared. The table 1 consists of the 20 features and 8 frames: the numbers of the frames calculated varies according to the speech signal length. The mean and the standard deviation for the complete LPC were calculated. The figure 4 shows the plot of LPC features for 8 frames of 0 to 9 for one person sample with Amplitude.

Table 1: Single person feature extracted from LPC for digit siffar se Nau (Zero to Nine)

Coefficients	F1	F2	F3	F4	F5	F6	F7	F8
0	2.0568	0.4481	1.5377	0.1420	1.6860	0.3192	1.0000	0.0157
0	1.8323	0.3816	1.1422	0.0946	1.5490	0.2125	1.0000	0.0387
1	1.7683	0.5782	1.0367	0.2248	1.4745	0.4231	1.0000	0.0001
1	0.2964	0.4617	0.4892	0.2125	1.1985	0.3648	1.0000	0.0691
2	1.6896	0.6135	1.0000	0.4354	1.3044	0.4907	0.9575	0.2055
2	1.7277	0.7376	1.2247	0.5584	1.4258	0.6396	1.0000	0.2656
3	1.3487	0.6403	0.4854	0.3309	0.3593	0.4409	1.0000	0.2915
3	1.0000	0.4170	0.7645	0.1710	0.9238	0.3831	0.5802	0.0525
4	1.0000	0.3610	0.5974	0.1117	0.7410	0.2780	0.4321	0.0422
4	1.0841	0.4923	0.6949	0.2998	1.0000	0.3902	0.1776	0.0504
5	1.0000	0.3753	0.6341	0.1682	0.9845	0.2381	0.1103	0.0642
5	4.1241	1.5716	2.8829	0.6507	3.9863	0.9645	2.8732	0.0901
6	4.0563	1.5510	3.0055	0.7242	4.0362	0.9942	2.0955	0.1638
6	1.0000	0.8463	2.6167	0.2967	3.1479	0.6850	2.4002	0.1078
7	1.1688	0.4696	0.4037	0.1531	0.5809	0.3296	1.0000	0.0999
7	3.7641	1.5299	1.3306	0.9307	3.3024	1.2484	2.7618	0.5611
8	3.0622	1.3673	2.5448	0.5027	1.0370	0.8500	2.3562	0.3285
8	2.7667	0.9958	2.0955	0.1915	2.2582	0.8251	1.7878	0.0769
9	3.1514	1.0010	2.5413	0.3844	2.9026	0.7417	1.6479	0.2436
9	0.9940	1.2770	2.3642	0.4263	2.6121	0.8616	2.1165	0.1368

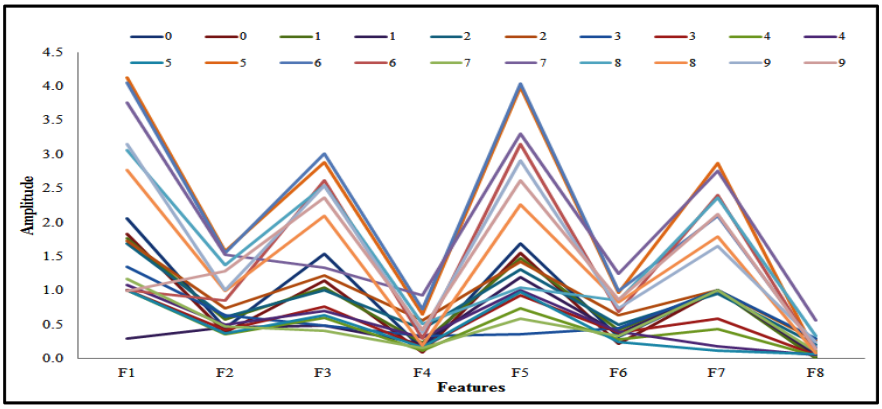

Figure 4: Plot of LPC features for 8 frames of 0 to 9 for one person sample

Table 2: Result for the numeric digits Zero to Nine (Sifer to Nau)

Sr. No.	Set of Numeric Digits	Female		Male		Average	
		PA %	WER %	PA %	WER %	PA %	WER %
1	Zero (Sifar)	80	20	100	00	90	10
2	One (Aik)	60	40	60	40	60	40
3	Two (Do)	80	20	60	40	70	30
4	Three (Teen)	80	20	80	20	80	20
5	Four (Char)	80	20	60	40	70	30
6	Five (Paanch)	80	20	60	40	70	30
7	Six (Chah)	100	00	40	60	70	30
8	Seven (Saat)	60	40	80	20	70	30
9	Eight (Aath)	80	20	60	40	70	30
10	Nine (Nau)	100	00	80	20	90	10
Average System PA and WER						**74**	**26**

Table 2 shows every digit PA and WER for Male and Female. Average System performance accuracy (PA) for Numeric data Zero to Nine (0-9) is 74% and WER is 26%. Number six and nine gives 100% result.

V. Conclusions

The efforts that are being carried out for the Urdu language are very less. After conducting the literature review we found that the language technologies can play a vital role in the development of good governing system. Apart from that we have used LPC for feature extraction and for classification HMM approach. Based on the result of this numeric recognition in Urdu language six and nine gives 100 percent result. The speech data collected in the noisy environment will help us to develop a robust speech recognition system.

References

[1] AbuZeina, D., & Elshafei, M. (2011). *Cross-Word Modeling for Arabic Speech Recognition*. Springer Science & Business Media.
[2] Akram, M. U., & Arif, M. (2004, December). Design of an urdu speech recognizer based upon acoustic phonetic modeling approach. In *Multitopic Conference, 2004. Proceedings of INMIC 2004. 8th International* (pp. 91-96). IEEE.
[3] M. A. Anusuya, S. K. Katti, "Speech Recognition by Machine: A Review", International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009, pp. 181-205
[4] Corneliu Octavian DUMITRU, Inge GAVAT, "A Comparative Study of Feature Extraction Methods Applied to Continuous Speech Recognition in Romanian Language", 48th International Symposium ELMAR-2006, 07-09 June 2006, Zadar, Croatia.
[5] DOUGLAS O'SHAUGHNESSY, "Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis", Proceedings of the IEEE, VOL. 91, NO. 9, September 2003, 0018-9219/03$17.00 © 2003 IEEE.
[6] N.Uma Maheswari, A.P.Kabilan, R.Venkatesh, "A Hybrid model of Neural Network Approach for Speaker independent Word Recognition", International Journal of Computer Theory and Engineering, Vol.2, No.6, December, 2010 1793-8201.
[7] G. 2003 Lalit R .Bahl et.al.,Estimating Hidden Markov Model Parameters so as to maximize speech recognition Accuracy,IEEE Transaction on Audio, Speech and Language Processing Vol.1 No.1 , Jan.1993.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Shaikh Naziya S. "LPC and HMM Performance Analysis for Speech Recognition Systemfor Urdu Digits." IOSR Journal of Computer Engineering (IOSR-JCE) 19.4 (2017): 14-18.