

# Linear Predictive Codes for Speech Recognition System at 121bps

Pramod B. Patil

*Principal, Dr. Rajendra Gode Institute of Technology & Research, Amravati, India  
ppamt07@yahoo.com*

## **Abstract**

*This paper described the recognition of the phonetics related to numerical in Indian regional languages such as Marathi & Hindi by Nearest Neighbour rule. The segmentation is based on the location of the start and endpoints of the speech. The exact speech boundaries can be located and evaluated for linear predictive codes. The Linear Predictive Codes of phonetics related to numerical in Indian regional languages such as Marathi & Hindi forms the codebook. The optimum distance between the test and the codebook linear predictive codes can be determined by the Dynamic Time Warping technique. Depending on the distance, the word is recognized by the Nearest Neighbour rule. The accuracy of 88% is achieved with a high reduction in the memory requirements & good SNR.*

**Keywords:** *Pitch, Linear predictive codes, Dynamic time warping, Nearest neighbor rule*

## **1. Introduction**

In nearly all speech recognition systems developed for any language, speech signals are preprocessed to extract the features such as pitch [1]. Speech signals in the time domain as raw data generated directly by the speech production system and contain all the acoustic information for recognition. The start and endpoints of the spoken speech are determined accurately. Two principal components of the speech production system, the vocal tract, and the excitation source can be parameterized by the time-varying autoregressive filter to find the Linear Predictive codes (LPC). The Speech nonstationary is represented in a compact and parametric form on a frame-by-frame basis. To implement optimization, the distance between frames of features is determined. The optimum distance between the LPC stored in the codebook and the currently spoken word is determined by the dynamic time warping technique.

## **2. System architecture**

Basic steps involved in the processing are:

1. End Point Analysis
2. Pitch Analysis
3. LPC Analysis
4. Pattern Classification
5. NN rule for the decision of recognition.

---

### **Article history:**

Received (December 31, 2019), Review Result (February 2, 2020), Accepted (March 4, 2020)

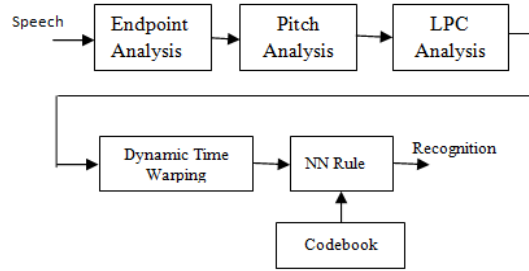


Figure 1. The Process diagram for phonetics related to numerical in Indian regional languages

The acoustic waveform of phonetics related to numerical in Indian regional languages such as Marathi & Hindi is analyzed for correctly detecting the start point, endpoint, and pitch of speech [2][3][4]. Linear Predictive coding analysis is performed to achieve the compressed form of large speech data. The optimum distance between the LPC's stored in the codebook and the spoken word under test respectively by the Dynamic time warping technique. The decision of recognition is performed by the Nearest - Neighbour rule.

### 3. Segmentation

The phonetics related to numerical in Indian regional languages such as Marathi & Hindi are sampled and analyzed for Linear Predictive Codes. The LPC's are stored to form the codebook. [Figure 2] shows the recorded original signal.

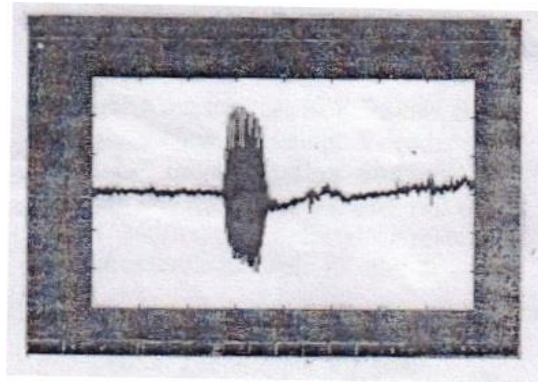


Figure 2. The recorded original signal

### 4. Pitch analysis

Pitch detection is an important task to represent speech for recognition. The autocorrelation technique estimates the pitch on a frame basis by using the sliding window technique to reduce the background noise [5][6][7][8]. The pitch correlation provided the sliding window is defined as:

$$R(T) = \max_{i=-T_s}^{T_s-1} [\max_T Ri(T)] \quad (1)$$

$$\text{where } Ri(T) = C(i, T + i) / \sqrt{C(i, i)C(T + i, T + i)} \quad (2)$$

$T_s$  - Maximum sliding range

$R_i(T)$  - Value of the normalized autocorrelation for the delay  $i$  Autocorrelation function

$$C_{n=0}^{N-1}(k, l) = \sum s(n+k)s(n+l) \quad (3)$$

where  $s(n)$  is the low pass speech signal,  $N$  is the frame size and  $k$  and  $l$  are the corresponding delays.

[Figure 3] shows the recorded original signal and the detection of pitch & endpoints for the phonetics related to numerical in Indian regional Marathi language //ek// respectively.

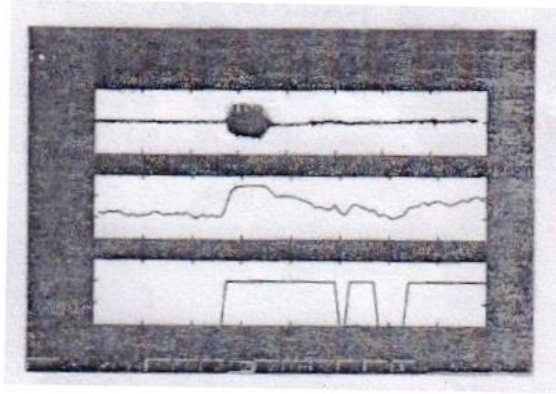


Figure 3. Detection of pitch & endpoints for the phonetics

## 5. LPC analysis

The speech sample can be approximated as a linear combination of past samples. By minimizing the sum of the squared differences over the frames between the actual samples and the linearly predicted ones, a set of predictive coefficients is determined [9][10][11][12][13].

For LPC analysis, hamming windowed samples of speech frames are processed by fixed order digital system to flatten the speech signal.

In linear prediction, the unknown output is represented as a linear sample, and the prediction coefficients are selected to minimize the mean square error.

If  $x_1, x_2, x_3 \dots x_n$  represent the data samples and  $x_{(n+1)}$  represent the predictor for the next sample  $x(n+1)$ , then

$$\begin{aligned} x(n+1) &= -(a_1x_n + a_2x_{n-1} + \dots + a_nx_1) \\ &= -\sum_{k=1}^n a_kx_{n+1-k} \end{aligned} \quad (4)$$

and the corresponding error

$$\begin{aligned} e(n+1) &\triangleq x(n+1) - \hat{x}(n+1) \\ &= x(n+1) - \sum_{k=0}^n a_kx_{n+1-k}, a_0 \triangleq 1 \end{aligned} \quad (5)$$

Minimization of the mean square error  $E[e^2(n+1)]$  concerning the unknown qualities  $a_1, a_2, \dots, a_n$  gives rise to the standard set of linear equations

$$= \sum_{k=0}^n a_{n-k}r_{k-i+1} + 0, i = 1 \rightarrow n.$$

$$= \sum_{k=0}^n a_{n-k}r_{k-n} \triangleq r^2$$

The matrix can be represented as

$$\begin{bmatrix} r_0 r_1 & \dots & \dots & \dots & \dots & r_n \\ r_1 r_0 & \dots & \dots & \dots & \dots & r_{n-1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_n r_{n-1} & \dots & \dots & \dots & \dots & r_0 \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ \dots \\ a_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ r^2 \end{bmatrix} \quad (6)$$

The unique solution

$$A_n(Z) = 1 + a_1 z^{-1} + \dots + a_n z^{-n} \quad (7)$$

subject to  $a_0=1$  represents the Levinsons polynomial of degree  $n$ . These polynomials have all zeros in  $|z| < 1$ , hence

$$H(Z) = \frac{r}{A_n(Z)} = \frac{r}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} \quad (8)$$

that generates  $x(n)$  representing a stable system. The LPC vectors  $[a_1, a_2, \dots, a_n]$  are generated to represent the speech. [Figure 4] shows the LPC spectrum.

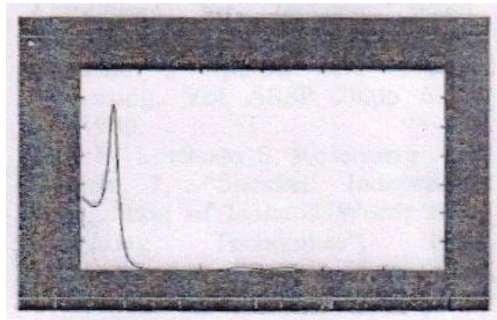


Figure 4. The LPC spectrum

## 6. Dynamic time warping framework

Once the patterns have been determined, the similarity between test and reference patterns is determined due to the highly variable speaking rate; pattern similarity involves both time alignment and distance computation performed simultaneously [14][15][16][17].

The alignment function  $w(t)$ , which maps reference pattern  $R$  onto the corresponding parts of test pattern  $T$  is measured by calculating, the optimized distance between the functions

$$d(T, R) = \|T - R\| = \sum_{i=0}^p (Ti - Ri)^2 \quad (9)$$

where  $Ti$  and  $Ri$  are the  $i_{th}$  components of the various vectors  $T$  and  $R$  respectively.

The Dynamic time warping (DTW) determines the optimum path, which minimizes the accumulated distance between the test and the reference patterns, subject to a set of the path and endpoint constraints.

## 7. Nearest Neighbor (NN) rule

The codebook contains  $R$  reference patterns  $R, i= 1,2, \dots, V$ , and for each pattern, the optimum distance is determined by the DTW technique. The NN rule is simply [17].

$i^* = \text{opt} [D^i]$  i.e. choose the pattern  $R^{i^*}$  with optimum distance as the recognized pattern.

## 8. Result

The phonetics related to numerical in Indian regional languages such as Marathi & Hindi are sampled at 8 kHz with 8 bits representing each sample. The data is processed in the frame of 200 samples by hamming window technique with 100-sample step size is selected for overlapping to smoothen the data to determine the vector of the pitch. The vector is obtained by pre-emphasizing speech data to flatten it using order digital filter  $H(z)=1-az^{-1}$  with coefficient  $a=-0.97$ . [Table 1] shows the start and end frames of phonetics related to numerical in Indian regional Marathi language //ek//. The performance of the speech recognition system using the NN rule is presented in [Table 2]. The requirement of memory for phonetics related to numerical in Indian regional Marathi language is summarized in [Table 3]. [Figure 2] shows the pitch boundaries for the phonetics related to numerical in Indian regional Marathi Language //ek// with system order 10.

Table 1. The Start & End frames of //ek//

Spoken Word= ek	Starting Frame	Ending Frame
Order= 10	38	58
Order= 10	48	68

Table 2. Performance of speech recognition System using NN rule

Speaker	Recognition Success
5	60
26	76.92
50	88

Table 3. Memory requirement

Numerical	Using Endpoint detection KBps	Using LPC Analysis KBps	SNR
//ek//	4.8	0.264	24
//don//	6.8	0.121	17
//teen//	5.2	0.286	26
//char//	4.4	0.033	22
//pach//	4.4	0.242	22
//saha//	3.8	0.209	19
//sat//	5.4	0.297	14
//aath//	6.4	0.352	16
//nau//	4.8	0.264	24
//shunya//	6.4	0.352	16

## 9. Conclusion

The phonetics related to numerical in Indian regional languages such as Marathi & Hindi are recognized by using Dynamic Time warping Technique and Nearest Neighbour rule very

effectively. The technique requires very little memory to increase the recognition speed with an appreciable signal-to-noise ratio.

## References

- [1] W.S. and Mada Sanjaya, "Speech recognition using Linear Predictive Coding (LPC) and Adaptive Neuro-Fuzzy (ANFIS) to Control 5 DoF Arm Robot," *Journal of Physics: Conference Series*, vol.1090, pp.1-10, (2018)
- [2] Oday Kamil Hamid, "Speech sound coding using linear predictive coding (LPC)," *Journal of Information, Communication, and Intelligence Systems (JICIS)*, vol.3, no.1, pp.13-17, (2017)
- [3] Samad S, Hussain A, and Fah L K, "Pitch detection of speech signals using the cross correlation technique," *Proceedings of IEEE on Speech, Audio, and Signal Processing*, pp.283-286, (2000) DOI: 10.1109/TENCON.2000.893673
- [4] Rabiner L R, and Sambur M R, "An algorithm for determining the endpoints of isolated utterances" *Bell System Technical Journal*, vol.54, pp.297-315, (1975)
- [5] M.J. Ross, H.L. Shaffer, A. Cohen, R. Freudberg, and H.J.Manley, "Average magnitude difference function pitch extractor," *IEEE Transaction on Acoustic, Speech and Signal Processing*, pp.353-362 Oct, (1974)
- [6] D. Takin, "A robust algorithm for pitch tracking (RAPT)," *Speech Coding and Synthesis*, Netherlands: Elsevier Science, (1995)
- [7] L.A. Atkinson, M. Kondo and B.G.Evans, "Time envelop vocoder, A new LP based coding strategy for use of bit-rate 2.4kb/s and below," *IEEE Journal on Selected Areas on Communications*, vol.13, no.2, pp.449-457, (1995) DOI: 10.1109/49.345890
- [8] B.Gold and L. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in the time domain," *The Journal of the Acoustical Society of America*, vol.46, pp.442-448, (1969) DOI: 10.1121/1.1911709
- [9] Douglas O. Shaughnessy, "Linear predictive coding one popular technique of analyzing certain physical signals," *IEEE Potentials*, pp.29-32, (1998)
- [10] Pillai S., Hyun S. Oh, and Akansu A., "A new parametric formulation for linear predictive coding," *Proceedings of IEEE on Signal Processing*, pp.1432-1435, (1995)
- [11] Kwong S. and Man K. F., "A speech coding algorithm based on predictive coding," *Proceedings of IEEE on Speech and Audio Processing*, pp.455-460, (1995)
- [12] Yakhnich E. and Bistriz Y., "Constant delay and rate coding of speech spectral envelope at 11 bits / frame," *Proceedings of IEEE on Speech, Audio and Signal Processing*, pp.247-229, (2002)
- [13] Paliwal K. and Atal B. S., "Efficient vector quantization of LPC parameters at 24 bits / frame," *IEEE Transaction on Speech and Audio Processing*, vol.1, no.1, pp.3-14, (1993)
- [14] Gray A. H. and Markel J., "Distance measures for speech processing," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol.24, pp.380-391, (1976)
- [15] Sakoe H. and Chiba S., "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol.26, pp.43-49, (1978)
- [16] Mayers C., Rabiner L. R., and Rosenberg A. E., "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol.28, pp.622-635, (1980)
- [17] L. Rabiner, S. Levinson, A. Rosenberg, and J. Wilpon, "Speaker independent recognition of isolated words using clustering techniques," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol.27, pp.336-349, (1979) DOI: 10.1109/TASSP.1979.1163259

## **Authors**

### **Dr. Pramod Patil**

working as Principal for the last 13 years and currently working as Principal at Dr. Rajendra Gode Institute of Technology & Research, Amravati, India. His total teaching experience is 30 years. He has supervised many research scholars for Ph.D. and Masters. He has published several papers in International Journal/ Conference. He is the author of two books on Image Processing and Speech Analysis.

Email id: ppamt07@yahoo.com, M# +917720002401

***This page is empty by intention.***