IASTED International Conference on Signal and Image Processing
(SIP 2006), August 14 - 16, 2006, at Honolulu, USA

LINEAR PREDICTIVE CODES FOR SPEECH RECOGNITION SYSTEM AT 121BPS

Pramod B. Patil Dr. Vijay T. Ingole Kanchan S Bhagat
MIEEE, MIETE SMIEEE, FIETE
PRM Institute of Technology and Research, Badnera, Amravati, India

Abstract:

This paper described the recognition of the phonetics related to numericals in Indian regional languages such as Marathi & Hindi by Nearest Neighbour rule. The segmentation is based on the location of the start and end points of the speech. The exact speech boundaries can be located and evaluated for linear predictive codes. The Linear Predictive Codes of the phonetics related to numericals 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 Nearest Neighbour rule. The accuracy of 88 % is achieved with highly reduction in the memory requirement& good SNR.

Keywords:

Pitch, Linear Predictive Codes, Dynamic Time Warping, Nearest Neighbour Rule

I. Introduction:

In nearly all speech recognition system developed for any language, speech signals are preprocessed to extract the features such as pitch [1]. Speech signal in time domain as raw data generated directly by the speech production system and contain all the acoustic information for recognition. The start and end points of the spoken speech is determined accurately. Two principle components of speech production system, the vocal track and the excitation source, can be parameterized by time varying autoregressive filter to find the Linear Predictive codes (LPC). The Speech nonstationary is represented in a compact and parametric form on frame-by-frame basis. In order 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 dynamic time warping technique.

II. 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 decision of recognition.

Fig 1. Shows the Process diagram for phonetics related to numericals in Indian regional languages

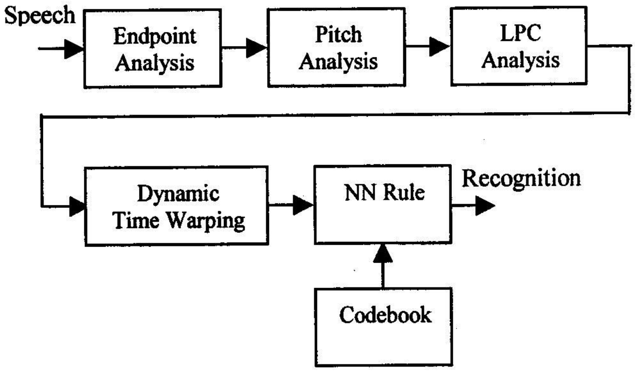

Fig 1. Process diagram for phonetics related to numericals in Indian regional languages

The acoustic waveform of phonetics related to numericals in Indian regional languages such as Marathi & Hindi is analyzed for correctly detecting the start point, end point and pitch of speech [2]-[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.

III. Segmentation:

The phonetics related to numericals 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.

IV. Pitch Analysis:

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

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

where Ri(T)

$= C(i, T+i)/\mathrm{sqrt}\{C(i,i)C(T+i,T+i)\}$

Ts – Maximum sliding range

Ri(T) - Value of the normalized autocorrelation for delay i

Autocorrelation function

$$C(k,l) = \sum_{n=0}^{N-1} s(n+k)s(n+l)$$

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

V. 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]-[13].

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

In linear prediction, the unknown output is represented as a linear combination of past known samples, and the prediction coefficients are selected to minimize the mean square error.

If x1, x2, x3 ....xn represent the data samples and $\hat{x}(n+1)$ represent the predictor for the next sample x(n+1) , then

$$x(n+1) = - (a_1x_n + a_2x_{n-1} + \ldots\ldots\ldots + a_nx_n)$$

$$= - \sum_{k=1}^{n} a_k\, x_{n+1-k}$$

and the corresponding error

$$e(n+1) \triangleq x(n+1) - \hat{x}(n+1)$$

$$= \sum_{k=0}^{n} a_k\, x_{n+1-k}\ ,\ a_0 \triangleq 1.$$

Minimization of the mean square error $E\ [e^2\ (n+1)]$ with respect to the unknown qualities a_1, a_2 ,........an 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

$$\begin{bmatrix} r_0\ r_1 \ \cdots\cdots\cdots\ r_n \\ r_1\ r_0 \cdots\cdots\cdots r_{n-1} \\ \cdots\cdots\cdots\cdots\cdots \\ r_n\ r_{n-1} \cdots\ \cdots\cdots r_0 \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ \\ a_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ r^2 \end{bmatrix}$$

The unique solution $A_n(z) = 1 + a_1 z^{-1} + \ldots\ldots + a_n z^{-n}$ subject to $a_0=1$ represents the Levinsons polynomial of degree n .These polynomials have all zeros in $| z | < 1$, hence $H(z) = r / A_n(z) = r / 1 + a_1 z^{-1} + \ldots\ldots + a_n z^{-n}$ that generates x(n) representing stable system. The LPC vectors $[a_1 , a_2, \ldots\ldots a_n]$ generated represents the speech.

VI. Dynamic Time Warping Framework:

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

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

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

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

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

VII. Nearest Neighbour (NN) Rule:

The codebook contains R reference patterns R^i , i = 1,2 ….V , and for each pattern the optimum distance is determined by 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.

VIII. Result:

The phonetics related to numericals in Indian regional languages such as Marathi & Hindi are sampled at 8KHz with 8 bits representing each sample. The data is processed in 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 pitch. The vector is obtained by pre-emphasizing speech data to flatten it using first 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 numericals in Indian regional Marathi language//ek//. The performance of speech recognition system using NN rule is presented in Table 2. The requirement of memory for phonetics related to numericals in Indian regional Marathi language is summarized in Table 3. Fig 2.shows the pitch boundaries for the phonetics related to numerical in Indian regional Marathi Language //ek// with system order 10.

Spoken Word= ek	Starting Frame	Ending Frame
Order=10	38	58
Order=11	48	68

Table 1.Showing the Start & End frames of //ek//

Speaker	Recognition success
5	60
26	76.92
50	88

Table 2. Performance of speech recognition System using NN rule

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

Table 3. Memory requirement

IX. Conclusion

The phonetics related to numericals in Indian regional languages such as Marathi & Hindi are recognized very effectively. The technique requires very less memory so as to increase the recognition speed with appreciable signal to noise ratio.

X. Figures

Figure 2& 3 shows the recorded original signal, and the detection of pitch & end points for the phonetics related to numerical in Indian regional Marathi language //ek// respectively.

Figure 4 & 5 shows the extraction of the original signal and zero crossing rate of the speech signal respectively.

Fig 6 are the LPC spectrum.

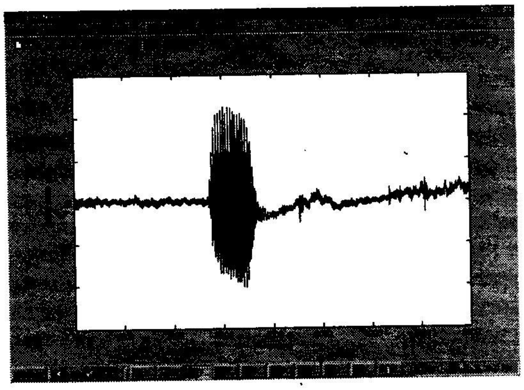

Figure2 shows the recorded original signal.

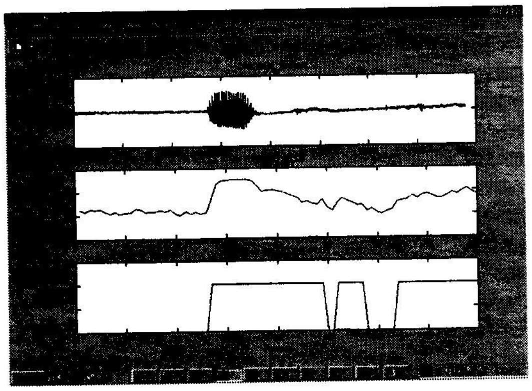

Figure3 the detection of pitch & end points for the phonetics.

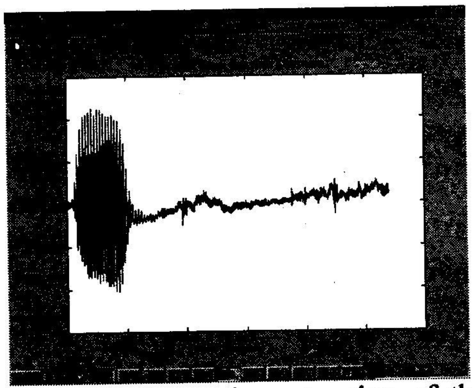

Figure 4 shows the extraction of the original signal.

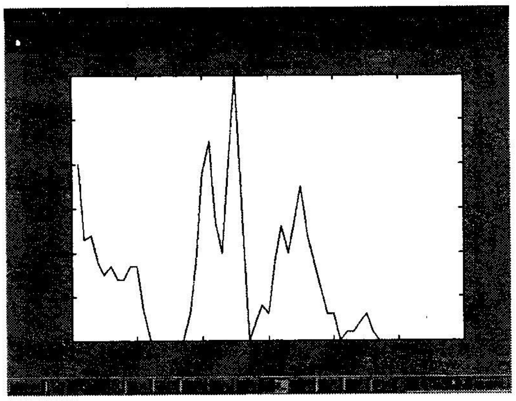

Figure 5 shows the zero crossing rate of the speech signal

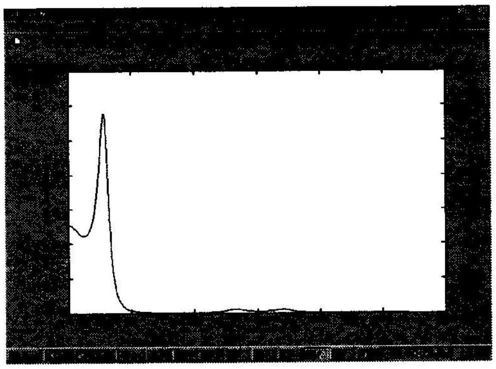

Figure 6 shows the LPC spectrum

XI. References:

1. Samad S, Hussain A, 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.

2.Rabiner L R, Sambur M R, "An Algorithm for Determining the Endpoints of Isolated Utterances" Bell System Technical Journal, Vol 54, pp 297 -315, 1975.

3.M.J.Ross, H.L.Shaffer, A. Cohen, R. Freudberg, and H.J. Manley, "Average Magnitude Difference Function Pitch Extractor", IEEE Trans.

4..Acoustic, Speech, and Signal Processing, pp. 353-362 Oct. 1974.

5. D. Takin, "A Robust Algorithm for Pitch Tracking (RAPT)", Speech Coding and Synthesis, Netherlands: Elsevier Science. 1995.

6. LA. Atkinson, M. Kondoz 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, Feb. 1995.

7. B.Gold and L. Rabiner, "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain", Journal of Acoustics. Society, America, Vol. 46, pp. 442-448, Aug. 1969.

8. Douglas 0 'Shaughnessy, "Linear Predictive Coding One Popular Technique of Analyzing Certain Physical Signals", IEEE Potentials, pp 29 – 32, Feb 1998.

9. Pillai S, Hyun S Oh, Akansu A,"A New parametric Formulation for Linear Predictive Coding", Proceedings of IEEE on Signal Processing, pp1432-1435, 1995.

10.Kwong S, Man K F, " A Speech Coding Algorithm Based on Predictive Coding", Proceedings of IEEE on Speech and Audio Processing, pp 455-460,1995.

11.Yakhnich E., Bistritz 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.

12.Paliwal K, Atal B S, "Efficient Vector Quantization of LPC Parameters at 24 Bits / Frame", IEEE Transaction on Speech and Audio Processing, Vol1, No1, pp 3-14, 1993.

13.Gray A H ,Markel J, “Distance Measures for Speech Processing”, IEEE Transactions on Acoustic, Speech and Signal Processing, Vol ASSP 24,pp 380-391,1976.

14.Sakoe H, Chiba S,”Dynamic Programming Algorithm Optimization for Spoken Word Recognition”, IEEE Transactions on Acoustic , Speech and Signal Processing, Vol ASSP 26,pp 43-49,1978.

15.Mayers C, Rabiner L R, Rosenberg A E , “Performance tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition”, IEEE Transactions on Acoustic , Speech and Signal Processing, Vol ASSP 28,pp 622-635,1980.

16.L R, Levinson S, Rosenberg A E, Wilpon J, “Speaker Independent Recognition of Isolated Words using Clustering Techniques”, IEEE Transactions on Acoustic , Speech and Signal Processing, Vol ASSP 27,pp 336-349,1979.