# Novel Protocol For Transmitting Linear Predictive Codes Of Speech On Computer Networks

# Pramod B. Patil[1] · V. T. Ingole[2] ,Kanchan S Bhagat[3] · Mahesh T. Kolte[4]
[1] PRM Institute of Technology and Research, Badnera, Amravati
pramod7568@yahoo.co.in
[2] PRM Institute of Technology and Research, Badnera, Amravati
vijayingole@hotmail.com
[3] J.T. Mahajan Cllege of Engineering, Faizpur
ksbhagat@indiatimes.com
[4] MHSS College of Engineering, Mumbai
mtkolte@yahoo.com

**Abstract**

This paper described the Novel Protocol for transmitting the packets of linear predictive codes of speech on computer networks. 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 LPCs of speech are transferred over the computer networks in a packet mode. 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 memory required is 1.1 K.

**Index Terms**

Linear Predictive Codes, Pitch, Protocol

## 1 Introduction

Fast development and wide distribution of computer networks have allowed using for organization of telephone. The global computer network internet enables to set telephone links between different countries The development of the system with acceptable quality requires speech compression algorithms adapted to Linear predictive codes (LPC) of speech over computer networks, speech flow transfer protocol and telephone network.
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).

## 2 LPC Analysis

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 recognition of phonetics related to numericals in Indian regional languages such as Marathi & Hindi.

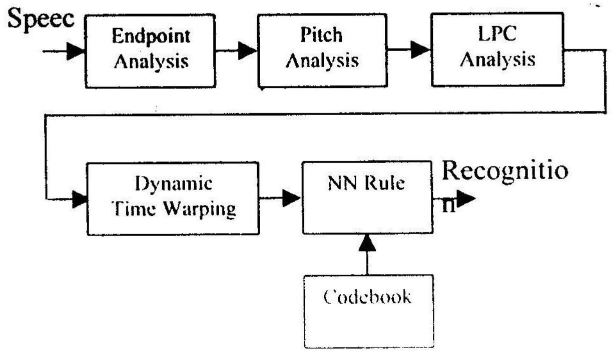


Fig 1. Shows the Process diagram for speech recognition.

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.

## 3 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.

## 4 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.

## 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]-[13].
For LPC analysis, hamming windowed samples of 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 x(n+1) represent the predictor for the next sample x(n+1), then

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

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

and the corresponding error

$$e(n+1) \overset{\Delta}{=} x(n+1) - \hat{x}(n+1)$$

$$= \sum_{k} a_k\, x_{n+1-k}\, , \; a_0 \overset{\Delta}{=} 1.$$

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

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

$$= \sum_{k=0}^{n} a_{n-k}\, r_{k}\;_{n} \overset{\Delta}{=} r^2$$

The matrix can be represented as

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

The unique solution
$A_n(z) = 1 + a_1z^{-1} + \dots\dots + a_nz^{-n}$ subject to $a_0=1$ represents the Levinsons polynomial of degree n. These polynomials have all zeros in $|z| < 1$, hence

$H(z) = r/A_n(z) = r/1 + a_1z^{-1} + \dots\dots + a_nz^{-n}$ that generates x(n) representing stable system. The LPC vectors $[a_1, a_2, \dots\dots a_n]$ generated represents the speech.

## 6 LPC of Speech Transmission over Computer Networks

The LPCs of speech are transferred over the computer networks in a packet mode. The packets delivery time can vary depend on the level of current network load. Thus each packet has unique delivery time that results in packets delay and their rearrangements in the flow on the reception side. It is necessary to minimize delay entering for restoring and to provide percent of lost speech packets not more than given threshold. Special rules of packets numbering are assigned to develop the new protocol. Defined 16 bit number is assigned for each LPC packet transmitted over the network. The rule of numbering is schematically represented in fig 2.

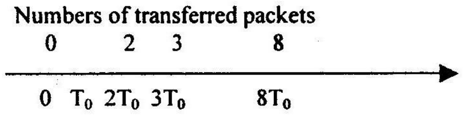


Fig 2 Packets numbering rule

The time axis is divided by equal intervals with $T_0$ duration. Each number of an upper scale coincides with the number of LPC of speech packet which has appeared at the encoder output in the given time moment. The packets are numbered inconsistently. The packet number corresponds to time interval when the packet was created. So in Fig first four packets in flow have numbers 0, 2, 3 and 8 accordingly. It is possible to restore the packets sequence in the flow and also to define the value of a time interval between adjacent packets by such numbering. These rules allow refusing of packet time stamp transmission, which is necessarily, is used in the existing protocols and also to considerably reduce the header size.

## 7 LPC s of Speech Packets Flow Restoring Algorithm

The main restoring service (RS) component is the buffer where the LPCs of speech packets received form the network are stored. Each packet has defined number. The frame number zero is fed to the decoder in the moment Ts. During the restoring process the next frame is selected from the buffer and in the moment TDi is transferred to the decoder. The moment of i frame decoding is defined by the formula
TDi = Ts + i. To .The a delay of restoring is not entered in the algorithm and the RS output flow structure coincides with transmitter output flow except for the late frames. The moment of decoding for packet i is needed for decoding is absent in the buffer at the moment the probable losses Lpr are estimated by the formula: Lpr = ($M_l$ /Mi). 100 %, where $M_l$ –quantity of lost packets and Mi – quantity of transferred packets. The packet i is assumed as lost. If these losses do not exceed the threshold Lmax , the skip is fixed . The defined adjustment of thresholds Lmax and Twmax can provide the steady work of the system in networks which state is stable and predictable. Flowchart 1 shows block scheme of the restoring algorithm.

## 8 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.

## 9 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 [17]
$i^*$ = opt[$D^i$] i.e. choose the pattern $R^{i*}$ with optimum distance as the recognized pattern.

## 10 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-a$z^{-1}$ with coefficient a= -0.97. Table 1 shows the start and end frames of phonetics related to numericals in Indian regional Marathi language//ek//. Fig 3 & 4 shows the pitch boundaries & the LPC spectrum for the phonetics related to numerical in Indian regional Marathi Language //ek// with system order 10.

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

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

Table 2 shows the memory requirement for storing the LPC s of speech signal.

| Numerical | Sampling At 8 KHz | Using Endpoint detection | Using LPC Analysis | SNR |
|---|---|---|---|---|
| //ek// | 64K | 38.4K | 2.1K | 1.2 |
| //don// | 64K | 54.4K | 2.9K | 5.6 |
| //teen// | 64K | 41.6K | 2.2K | 4.3 |
| //char// | 64K | 35.2K | 1.9K | 4.3 |
| //pach// | 64K | 35.2K | 1.9K | 1.1 |
| //saha// | 64K | 30.4K | 1.6K | 0.86 |
| //sat// | 64K | 43.2K | 2.3K | 2.25 |
| //aath// | 64K | 51.2K | 2.8K | 1.28 |
| //nau// | 64K | 38.9K | 2.1K | 0.75 |
| //shunya// | 64K | 51.2K | 2.8K | 1.6 |

Table 2. Memory requirement

## 11 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.
The necessity of new transfer protocol development for LPCs of real time speech transmission over computer networks in Internet telephony is considered.

### Figures & Flowchart

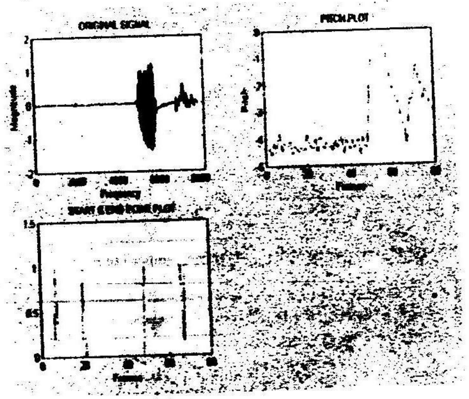


Fig 3 pitch boundaries of speech signal

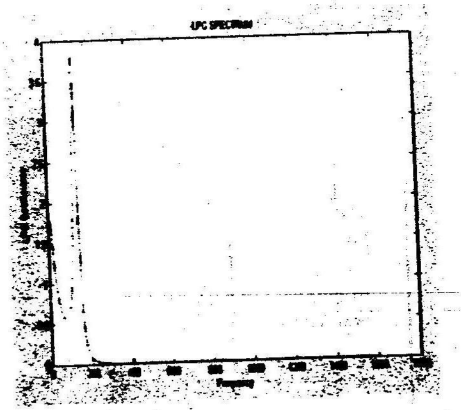
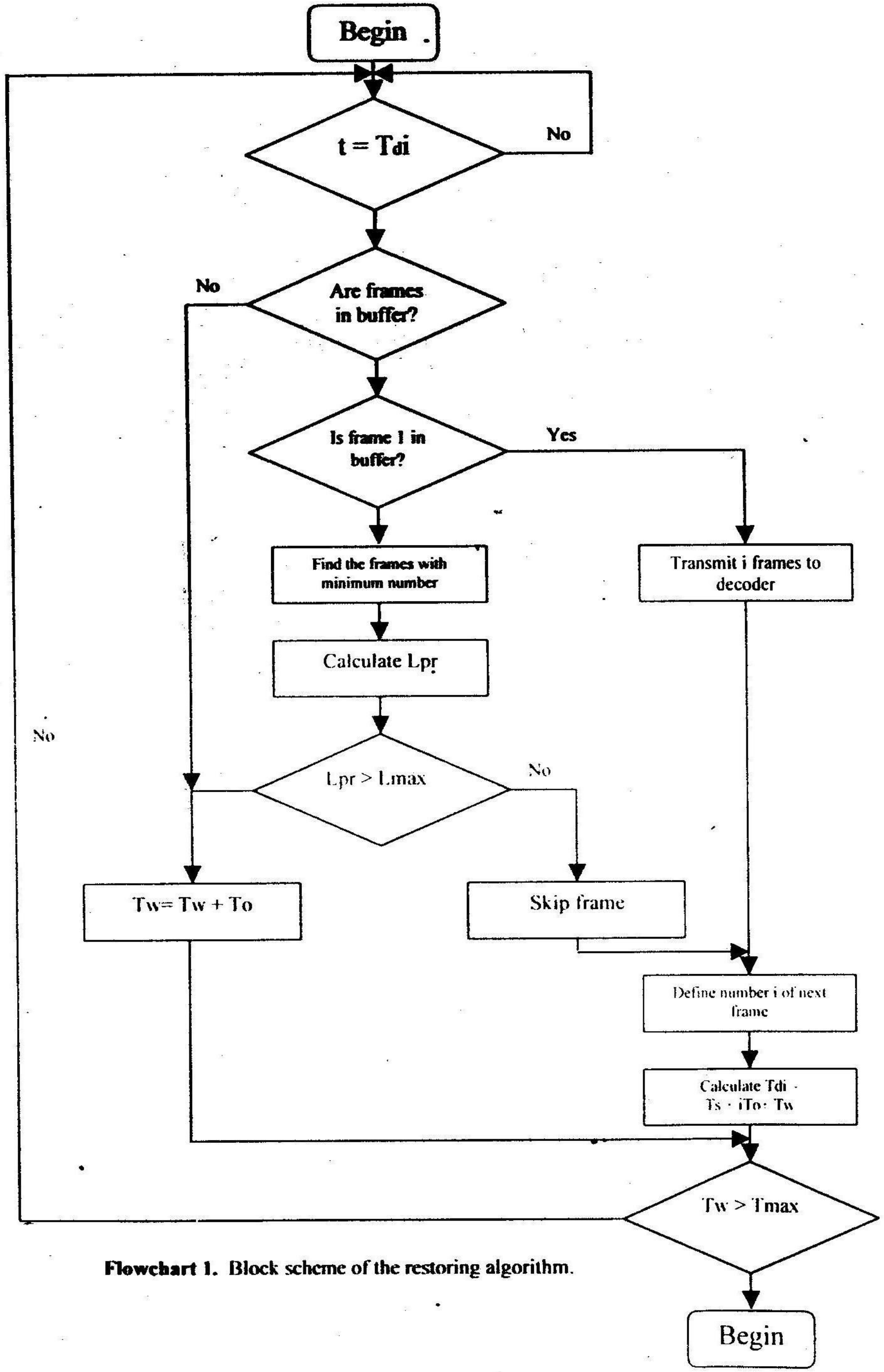


Fig 4 shows the spectrum of LPC.



**Flowchart 1.** Block scheme of the restoring algorithm.

## 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. Levinson S E., "Isolated and Connected Word Recognition – Theory and Selected Applications ". *IEEE Transaction on Communications*, Vol 29. No 5. 1981

[3] . Rabiner L R. Sambur M R. " Speaker Independent Recognition Of Connected Digits". *Bell System Technical Journal* Vol 54. pp 202 –205. 1972.

[4] 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.

[5] M.J.Ross. H.L.Shaffer. A. Cohen. R. Freudberg. and H.J. Manley. "Average Magnitude Difference Function Pitch Extractor". *IEEE Trans. Acoustic. Speech. and Signal Proc*. pp. 353-362 Oct. 1974.

[6] D. Takin. "A Robust Algorithm for Pitch Tracking (RAPT)". *Speech Coding and Synthesis. Netherlands: Elsevier Science*. 1995.

[7] 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.

[8] 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.

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

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

[11] 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.

[12] 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.

[13] 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.

[14] Postel J B. " Transmission Control Protocol". *RFC 793*.1981.

[15] Postel J B. "Internet Protocol". *RFC 791*.1981.

[16] 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.

[17] Rabiner 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.