ISCA Journal of Engineering Sciences ______________________________ *ISCA J. Engineering Sci.*
Vol. **1(1),** 57-61, July **(2012)**

Vocoder (LPC) Analysis by Variation of Input Parameters and Signals

Gupta Rajani, Mehta Alok K. and Tiwari Vebhav
Truba College of Engineering and Technology Indore, INDIA

Available online at: www.isca.in
Received 31st October 2011, revised 29th January 2012, accepted 30th March 2012

Abstract

In this paper we have performed analysis on the quality of processed signals by varying various parameters associated with the Linear Predictive Coder (LPC) vocoder, designed as a project. The aim of the project was to compress (encode) voice signals using a lossy compression technique called LPC and obtain a compression rate up to 95% to utilize channel bandwidth by using less resources. But while performing the analysis this time we have given priority to the quality compromising a little on the efficiency of the process. In this paper we analyze the effect of sampling rate, order of the vocoder and size of the frame on standard male and female voice signal patterns. The LPC vocoder is designed using MATLAB®.

Key words: LPC, V, UV, u (n), s (n).

Introduction

Linear predictive coding (LPC) [1] is a means, used generally in audio signal processing and speech processing. It is used for representing the spectral envelope of a digital signal of speech in condensed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate. It provides extremely precise estimates of speech parameters.

LPC starts with the assumption that a speech signal is produced by a buzzer at the end of a tube (representing voiced sounds), with occasional added hissing and popping sounds. Although apparently rudimentary, this model is actually a close approximation of the reality of speech production. LPC analyzes the speech signal[2,3] by estimating the formants, thus removing their effects from the speech signal and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the left over signal after the subtraction of the filtered modeled signal is called the residue.

Material and Methods

There are a number of methods used to implement LPC. Out of these, three methods are primarily used, namely: i. The autocorrelation method using Levinson Durbin Algorithm[4]. ii. The covariance method[5]. iii. The lattice method[6]. Table-1 shows the comparison of these methods and after analysis we have used the autocorrelation method because of the its associated advantages.

Mathematical Model: The figure-1 and figure-2 show the general model of the speech generation and their mathematical model for the analysis respectively.

S(n) speech samples are related to the excitation u(n) by the simple difference equation.

$$s(n) = \sum_{k=1}^{p} a_k\, s(n-k) + Gu(n) \tag{1}$$

A linear predictor with prediction coefficients[7] $\propto_k$ is defined as a system whose output is

$$\bar{s}(n) = \sum_{k=1}^{p} \propto_k s(n-k) \tag{2}$$

The system function of a pth order linear predictor[8] is the polynomial

$$P(z) = \sum_{k=1}^{p} \propto_k z^{-k} \tag{3}$$

The prediction error e(n), is defined as

$$\begin{aligned} e(n) &= s(n) - \bar{s}(n) \\ &= s(n) - \sum_{k=1}^{p} \propto_k s(n-k) \end{aligned} \tag{4}$$

It can be seen that the prediction error sequence is the output of a system whose transfer function is given by

$$A(z) = 1 - \sum_{k=1}^{p} \propto_k z^{-k} \tag{5}$$

Thus the prediction error filter A(z) will be an inverse filter for the system H(z) i.e.

$$H(z) = \frac{G}{A(z)} \tag{6}$$

Parameters: The following parameters are used to perform analysis: i. Male/ Female Voice pattern. ii. Sampling Frequency (Fs). iii. Order of the Vocoder (L). iv. Size of Time Frame, in ms (fr)[9]. v. Size of the window, in ms (fs). The results of these analyses are presented below in the same order.

The analysis was performed on three standard voice samples of both the sexes spoken in English (American Accent). The default settings coded in the vocoder[10] were used.

Male/Female Voice patterns: In the analysis it was found that the vocoder had better response for the voice samples of the fairer sex. The decoded voice signal obtained for the male voice pattern was distorted to the level of misinterpretation. This can be attributed to the low pitch of the male voice as compared to that of the females. Pitch attributes to the high amplitude peaks in a speech signals.

Sampling Frequency (Fs): The sampling frequency was varied (test file was a female voice sample). It was observed that a sampling frequency of less than 8000Hz degraded the performance as predicted by Nyquist's criterion. As we go on increasing the sampling frequency the quality of the output of decoder improves but only up to a limit. Here figure-3 shows the input signal, figure-4 and figure-5 show the decoded output at Fs be 8 KHz and 12 KHz resp. We also analyzed that after a certain frequency (varies from signal to signal) the quality again goes on debasing till the signal becomes incoherent.

Order of the Vocoder (L): As we go on increasing the order, which are the number of predictor coefficients[4] (number of poles of the output filter) the response of the system improves with an increase in depth of sound as shown in figure-6 and figure-7 respectively. But this quality comes with a price, the drawbacks being the consumption of more memory by variables and a slight echo effect.

Size of Time Frame, in ms (fr): Size of 'fr' decides speed (rate) at which the input samples are read and output sound is reproduced. Lowering the 'fr' results in slow playback and increasing the 'fr' in fast playback.

Size of Window, in ms (fs): If we increase the size of 'fs' then the frames start getting overlapped and a mixed sound (due to simultaneous playback/storage of decoded data of different time frames 'fr' on the same window 'fs') is heard as shown in figure-8 and figure-9 respectively

Note: 'fs' cannot be kept less than 'fr' as this will lead to a loss of bits in the output data frame.

Results and Discussion

We have tried to calculate the effective range at which the designed vocoder gives the best output together with an efficient use of memory. The optimum values of the analyzed parameters are: Fs = 10 KHz – 35 KHz, L = 13 – 25, fr = 20ms, fs = 30ms - 35ms. The vocoder works best with female voice samples with a lower order and in case of male voice samples it requires a higher order.

Conclusion

Linear Predictive Coding is an analysis/synthesis technique to lossy speech compression that attempts to model the human production of sound instead of transmitting the sound wave. Linear predictive coding achieves a compression bit rate of 2400 bits/second (MAXIMUM LIMIT) which makes it ideal for use in secure telephone systems. Secure telephone systems are more concerned that the content and meaning of speech, rather than the quality of speech, be preserved. The trade off for LPC's low bit rate is that it does have some difficulty with certain sounds and it produces speech that sound synthetic.

Linear predictive coding encoders break up a sound signal into different segments and then send information on each segment to the decoder. The encoder send information on whether the segment is voiced or unvoiced and the pitch period for voiced segment which is used to create an excitement signal in the decoder. The encoder also sends information about the vocal tract which is used to build a filter on the decoder side which when given the excitement signal as input and reproduce the original speech.

References

1. Rabiner L.R. and Schafer R.W., Digital Processing of Speech Signals, *Prentice-Hall Englewoods Cliffs,* Chapter 3, 38-103, Chapter 5, 172-232, Chapter 8, 396-453 (**1978**)
2. Sosnin I.N. and Rajfeld M.A., Algorithm of compensation of an acoustic noise for improvement of operation of digital algorithms of a speech coding (vocoders), *Electrical and Computer Engineering*, 146-149 **(2010)**
3. Cao Binshi, Sub band synthesized LPC vector quantization (SBS-LPC-VQ), *IEEE Workshop*, 114-116 **(2000)**
4. Harish D. and Ramasubramanian V., Comparison of segment quantizers: VQ, MQ, VLSQ and unit-selection algorithms for ultra low bit-rate speech coding, *Acoustics, Speech and Signal Processing*, 4773-4776 **(2008)**
5. Nascimento F.A.R. and Fraga F.J., New methods for improvement of sinusoidal transform vocoders, *ICME '04*, **2**, 1159 - 1162 **(2004)**
6. Makinen J., Ojala P. and Toukomaa H., Performance comparison of source controlled GSM, AMR and SMV Vocoders, *Intelligent Signal Processing and Communication Systems*, 151-154 **(2004)**
7. Jurafsky Daniel and James H. Martin, Speech and language processing, *Prentice Hall* **(2000)**
8. Yallapragada R. and Kripalani V., Increments in voice capacity and impact on voice quality with new vocoders in GSM and CDMA systems, *IEEE International Conference*, 100-104 **(2002)**
9. Sosnin I.N. and Rajfeld M.A., Algorithm of compensation of an acoustic noise for improvement of operation of digital algorithms of a speech coding (vocoders), *Electron Devices and Materials Proceedings,* 146- 149 **(2003)**
10. Islam M.S., Islam M.N., Alam M.S., Riaz M.A. and Hasan M.T., Performance evaluation of various vocoders in mobile ad hoc network (MANET), *Electrical and Computer Engineering*, 670-673 **(2010)**

Table-1
Comparison of Different Methods

Parameters	Amount of Storage	Number of Multiplication	Stability	Parker's coefficient	N
Autocorrelation	√	√	√	√	√
Covariance	√	-	-	√	√
Lattice	-	-	√	√	√

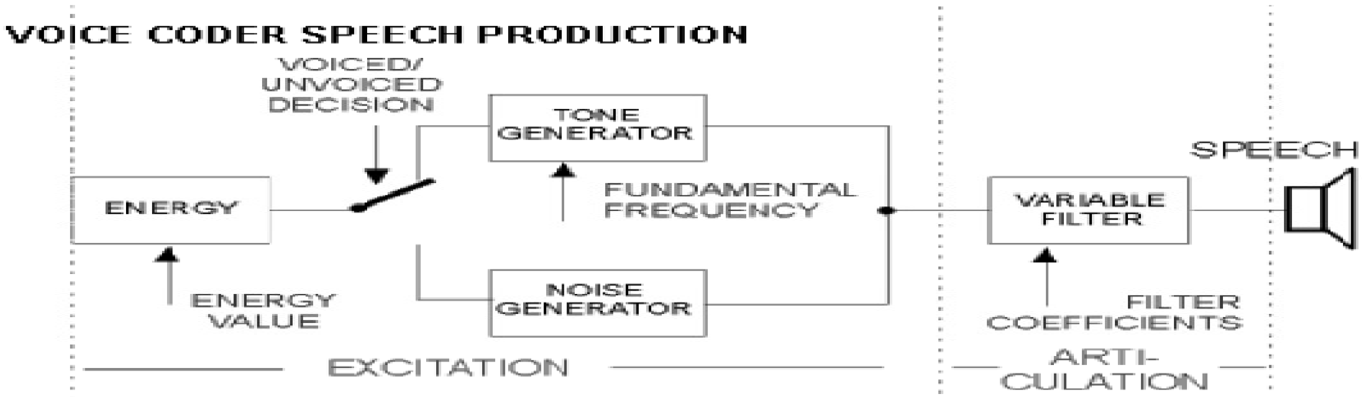

Figure-1
General Voice coder speech production

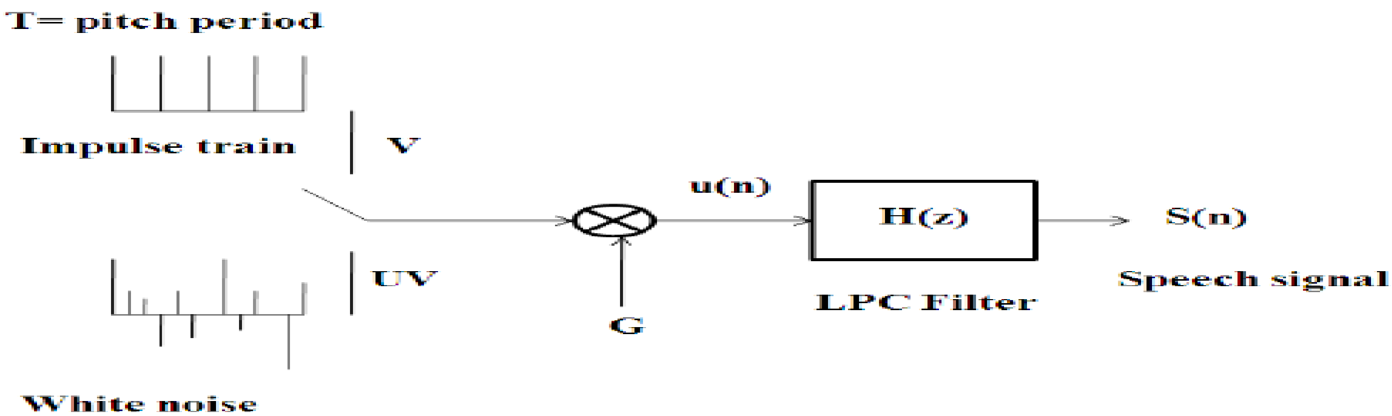

Figure-2
Mathematical Model

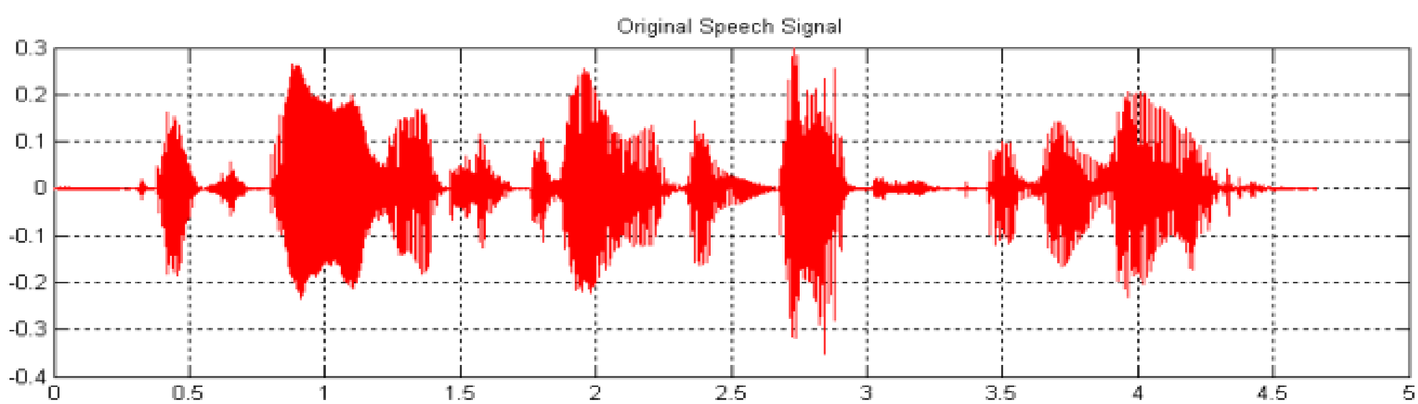

Figure-3
Input signal

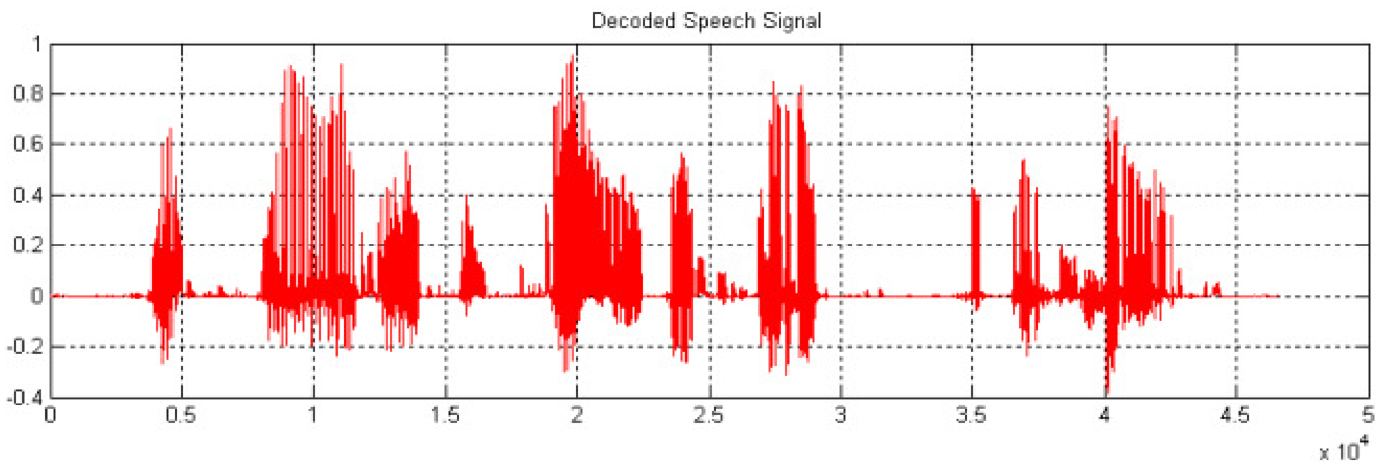

Figure-4
Speech signal decoded by sampling frequency 8 KHz

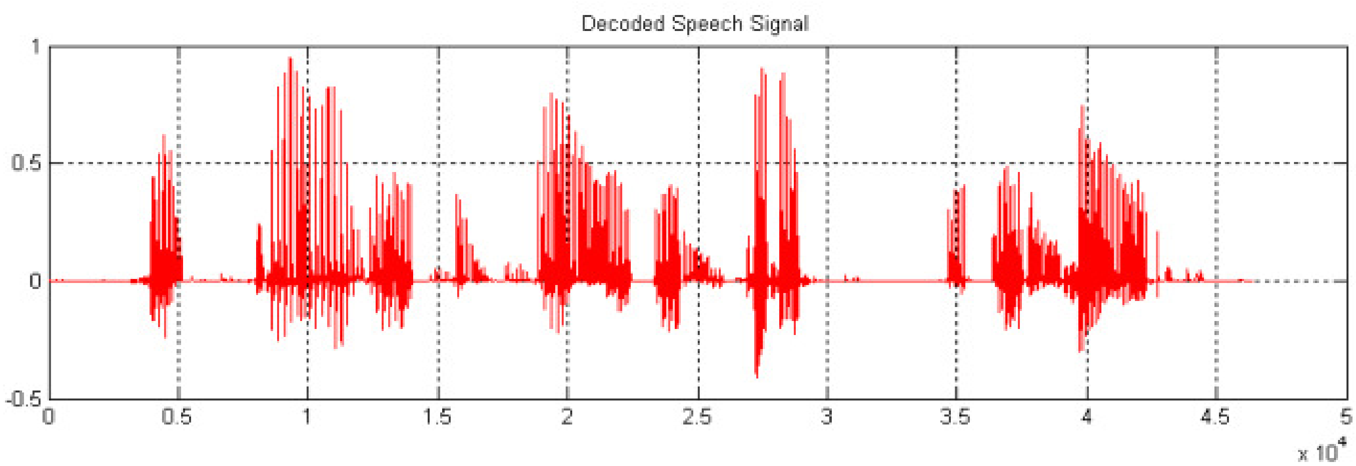

Figure-5
Speech signal decoded by sampling frequency 12 KHz

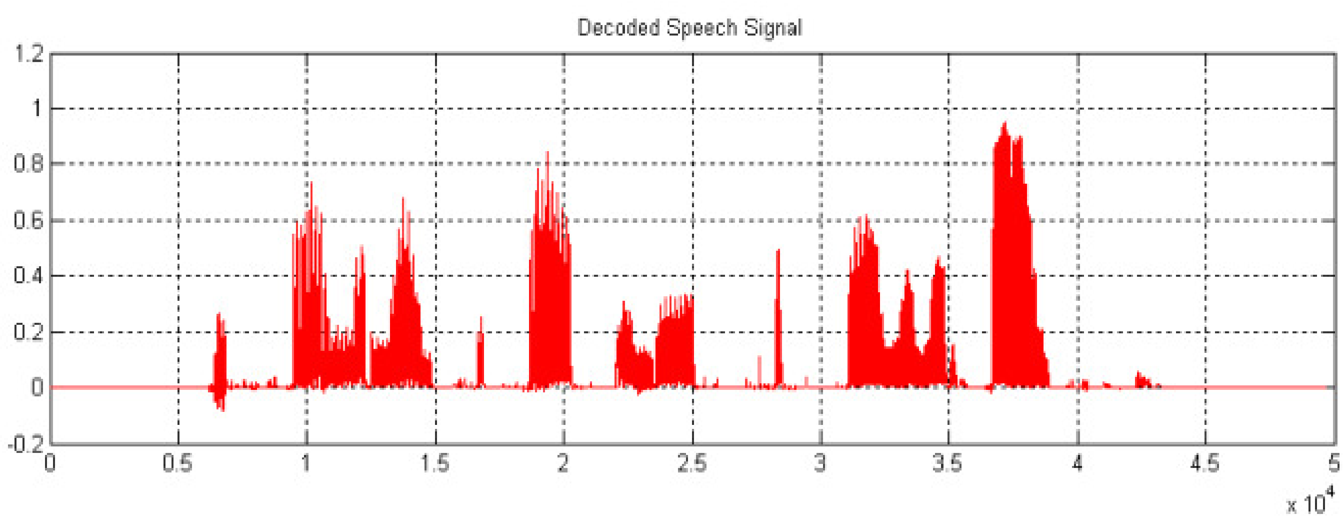

Figure-6
Speech signal decoded by 13 order vocoder

Vol. **1(1),** 57-61, July **(2012)**

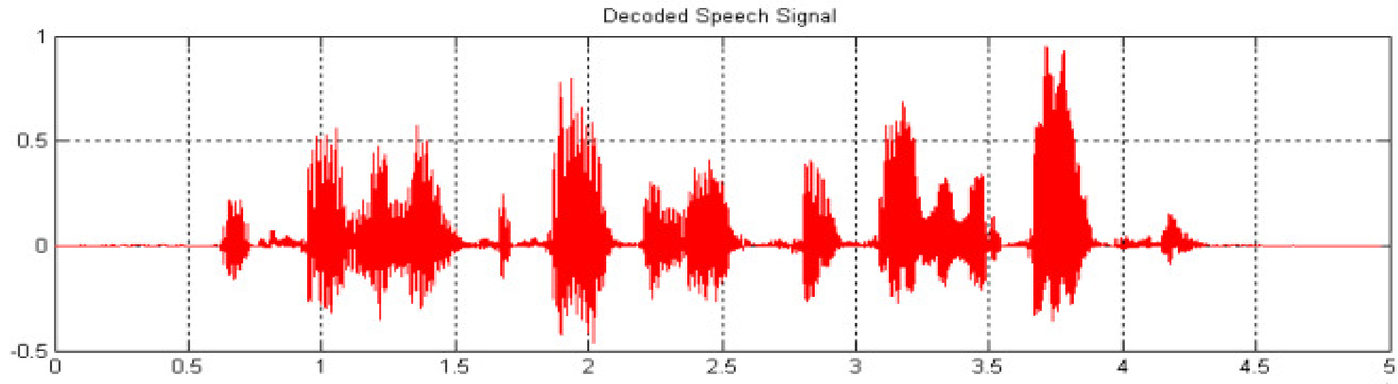

Figure-7
Speech signal decoded by 120 order Vocoder

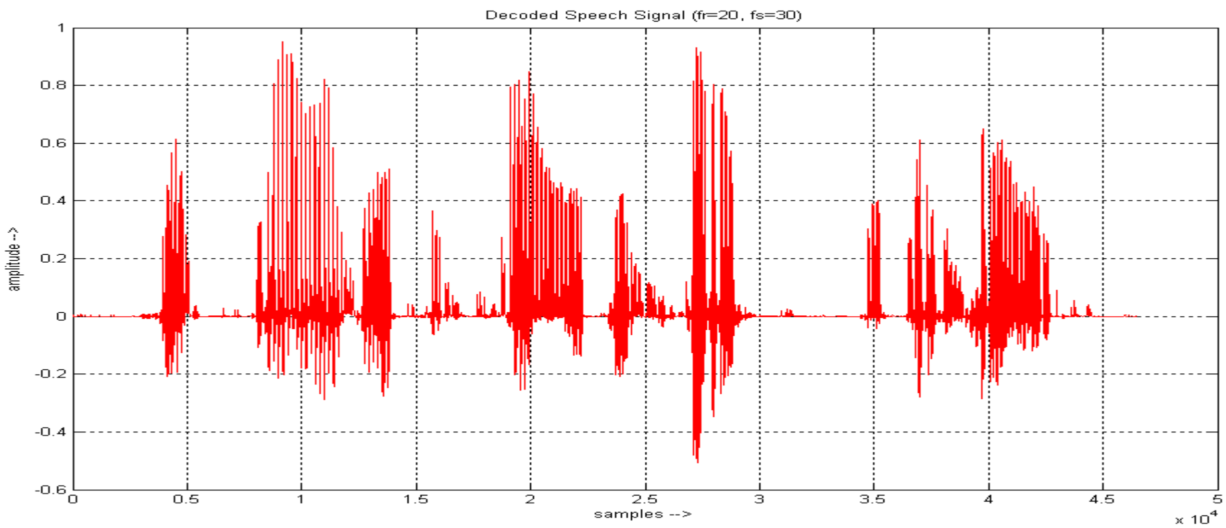

Figure-8 Speech signal decoded by 30ms window size

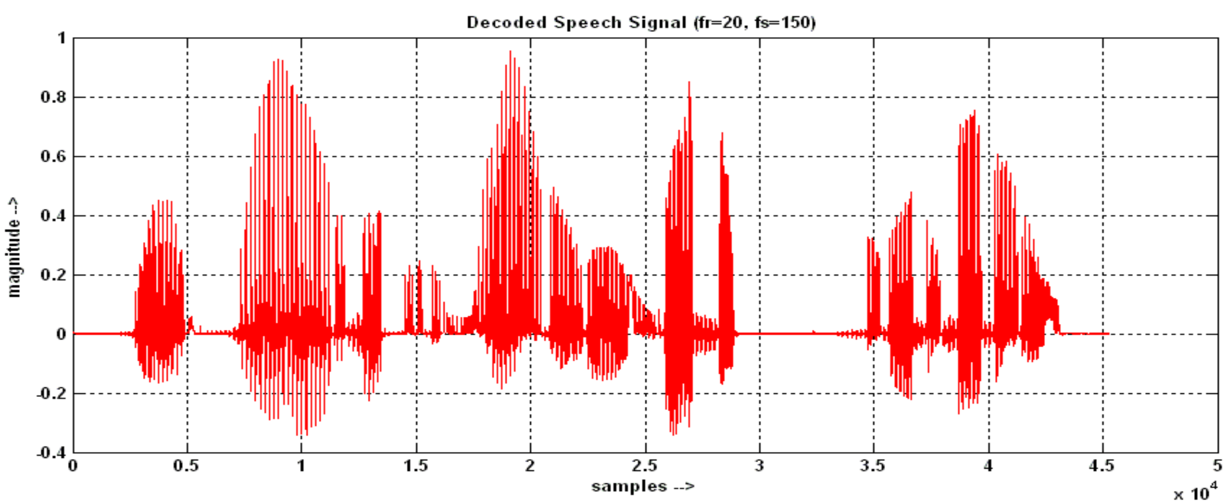

Figure-9
Speech signal decoded by 150ms window size