Military Technical College
Kobry El-Kobbah,
Cairo, Egypt

6th **International Conference**
on Electrical Engineering
ICEENG 2008

ICEENG

A new technique for regular pulse predictive coding of speech at low bit rates

By

Ibrahim M. Mansour
imansour@ju.edu.jo*

Samer J. AL-Abed
eng_samer1@yahoo.com*

Ahmad. K. Atieh
aatieh@btiphotonics.com**

Abstract:

Speech coding is a very important area that finds civilian and military applications. It can be considered as one of the important stages in speech processing. It is used to compress speech; this is because the speech signal is very redundant. Speech coding has many applications; it is used in digital telephony, in multimedia and in security of digital communications. In this paper, we focused on developing algorithms and methods for a waveform speech coder operating at low bit rate with good quality reconstructed speech signal. Moreover, a new model for linear predictive coding of speech that can be used to produce high quality speech at low data rate is introduced. In this model, we divided the residual (excitation signal) to subframes and made energy and voice / unvoice classifications to choose the best pulses in the residual that give us low bit rate and good quality for the reconstructed speech. Hence, this vocoder forms an excitation sequence which consists of groups of uniformly spaced pulses. During analysis the amplitude and LP coefficients of the pulses are determined. In addition, a new technique in the quantization of the amplitude of each pulse as well as linear prediction parameters is proposed.

Keywords:

Speech coding; Regular Pulse Excitation; LPC; speech compression; waveform speech coder

_____

 * Department of Electrical Engineering, Faculty of Engineering, University of Jordan, Amman 11942 Jordan Tel.: 00 962 7 77498906, Fax: 00 962 6 5355588
 ** BTI Photonic Systems Inc, Ottawa, Ontario, Canada K2H 5Z6

1. Introduction:

Speech coding is an important aspect of modern telecommunications. It deals with the problem of representing the speech signal with the fewest number of bits, while maintaining a good quality of the reconstructed speech with reasonable computational complexity and small delay. To accomplish high quality reconstructed speech at a low bit rate, coding algorithms are applied to remove the redundant information from the speech signal and so a smaller bit rate is required for transmission.

Wired communications have very large bandwidths as a result of the introduction of optical fiber, but in wireless and satellite communications bandwidth is limited. On the other hand, multimedia communications and other applications need to store the digitized voice. Reducing the data rate implies that less memory is needed for storage. These applications of speech compression make speech coding an attractive field of research.

The encoder (analysis stage) and the decoder (synthesis stage) are the two main components of a speech coder. The encoder encodes the speech signal in a compact form using fewer parameters. The decoder reconstructs the speech signal from those transmitted parameters. The analog speech signal $s(t)$ is first sampled at rate $f_s \geq 2f_{max}$, where f_{max} is the maximum frequency content of $s(t)$. The sampled discrete time signal is denoted by $s(n)$. After that many coding schemes such as PCM (pulse code modulation) or predictive coding are used to encode the signal $s(n)$.

In PCM (pulse code modulation) coding, the discrete time signal $s(n)$ is quantized to one of the 2^R levels. Each sample $s(n)$ is represented by R bits. At the decoder, these binary bits are used to reconstruct the speech signal. On the other hand, in predictive coding, the encoder takes a group of samples at a time, extracts parameters from these samples, and then converts those parameters to binary bits. At the decoder, these parameters are used to reconstruct the speech signal.

Linear Predictive Coding (LPC) method is one of the most important speech coding techniques for low bit rate speech coding. Since it has been so widely studied and applied [1] – [12]. It was initially developed in the late 1960's. Linear Predictive Coding (LPC) is a type of speech coding that can predict the signal from the past samples depends on the autocorrelation function [13], [14]. LPC removes the redundancies of a speech signal by modeling the speech signal as a linear all-pole filter, excited by a signal called the excitation signal (residual signal). Speech coders process a certain group of samples called a frame or a segment. In other words, Linear Predictive Coding (LPC) method for speech analysis and synthesis is based on modeling the vocal tract by a time varying all poles model and using it to remove a lot of redundancy in the speech signal [14], [15].

The speech encoder at the transmitter must determine the filter coefficients and the

excitation signal for each frame. The linear predictive coefficients (the filter coefficients) are selected to minimize the energy of the residual signal at the output of the filter for that frame [13], [16]. This filter is called a linear prediction analysis filter. The speech signal is first filtered through the linear prediction analysis filter. The resulting signal is called the residual signal for that particular frame. At the decoder, the inverse of the linear prediction analysis filter works as the linear prediction synthesis filter, while the residual signal acts as the excitation signal for the linear prediction synthesis filter. In order to reduce the total bit rate, speech coders such as LPC-10 (Linear Predictive Coding with a 10th – order model) do not transmit the whole residual signal [14], [15], [17]. In practical systems, the speech encoder at the transmitter must determine the filter coefficients and the proper excitation signal for each frame. To summarize, the whole model can be decomposed into the following two parts, the analysis part as shown in Fig. 1 (a) and the synthesis part as shown in Fig. 1 (b). The encoder (analysis part) analyzes the speech signal and produces the error signal. At the receiver, the decoder (synthesis part) takes the error signal as an input. The input is filtered by the synthesis filter $1/A(z)$, and the output is the speech signal. The error signal $e(n)$ is sometimes called the residual signal or the excitation signal. If the error signal from the analysis part is not used in synthesis, or if the synthesis filter is not exactly the inverse of the analysis filter, the synthesized speech signal will not be the same as the original signal. To differentiate between the two signals, we use the notation $s'(n)$ for the synthesized speech signal.
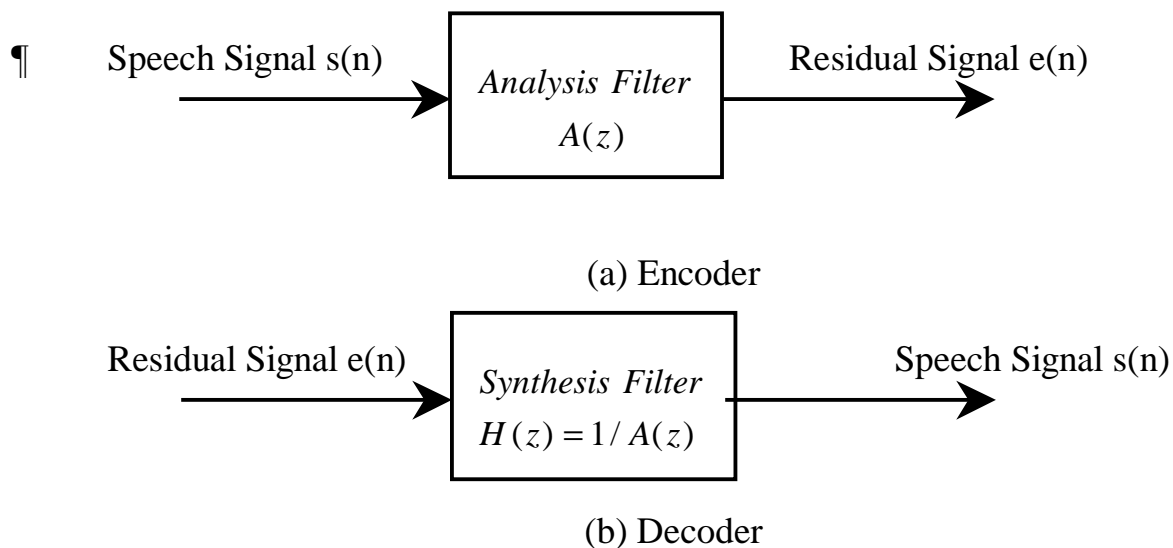
¶   Speech Signal s(n)   **Analysis Filter** $A(z)$ Residual Signal e(n)

(a) Encoder

Residual Signal e(n)   **Synthesis Filter** $H(z) = 1/A(z)$ Speech Signal s(n)

(b) Decoder

Figure 1: Linear prediction analysis and synthesis model (a) Encoder, (b) Decoder

2. Source Filter Model:

Speech signals are non-stationary, and at best they can be considered as stationary over short segments (typically 5-30 ms). Hence, the human speech production process is characterized by two factors: the source excitation and the vocal tract shape. In order to model speech production we have to model these two factors [11], [14], [15], [17]. The vocal tract is modeled as an all-pole transfer function H(z).

The vocal tract model $H(z)$ is excited by a discrete time glottal excitation signal $e(n)$ to produce the speech signal $s(n)$.

3. Analysis-Synthesis Model:

The ideal excitation for linear predictive coding synthesis is the prediction residual (the difference between estimated and exact signal):

$$e(n) = s(n) - \tilde{s}(n)$$

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

In classical linear predictive coding, this excitation is modeled by a pitch periodic impulse sequence for voiced speech and a random noise sequence for unvoiced speech [14], [15], [16], [17]. A different approach is taken in Residual Excited Linear Prediction (RELP), Multi-Pulse Linear Prediction (MPLP) and Regular-Pulse Excitation (RPE) [14], [15].
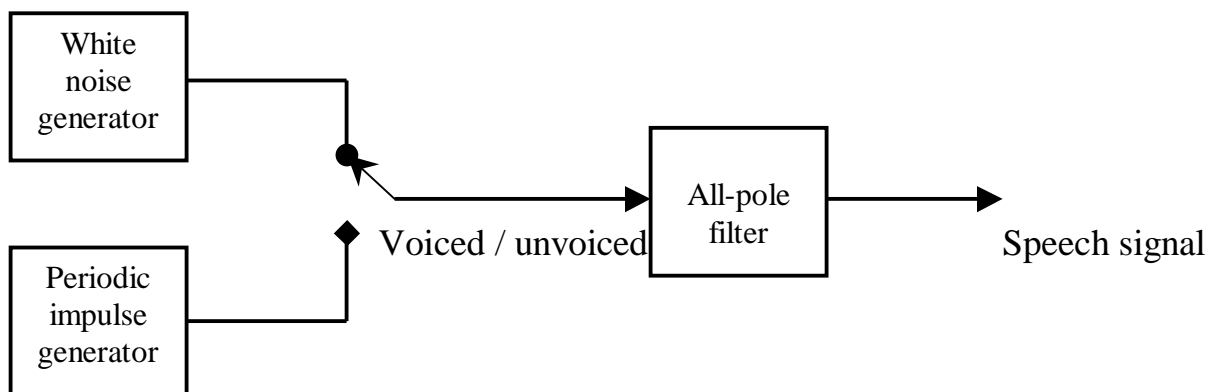
¶

Figure 2: A block diagram shows the source filter model of speech production

4. The Encoder Stage:

In our new model, we intend to improve the linear predictive model to reduce the bit rate of the speech signal while maintaining the perceptual quality of the reconstructed speech. In this model, we use the linear predictive coding method with a 10th order model for the voiced frame and an 8th order model is used for the unvoiced frame. The speech is sampled at 8 kHz to generate the linear prediction parameters for each frame. We divide the original speech to main frames. Each of these main frames is filtered to get the residual signal that is also divided to 6 subframes. We make energy classification to the subframes so that the subframe with higher energy will get more pulses than the subframe with lower energy. In addition to that, we make another classification step that classifies the main frames to voiced and unvoiced frames, so that the unvoiced frame will have reduction in the pulses with respect to the voiced frames. The purpose of dividing the residual to 6 subframes and making the energy and voice / unvoice classifications is to choose the best pulses in these subframes that give us low bit rate and good quality for the reconstructed speech.

5 Encoder Block Diagram:

At the encoder, linear prediction is used for the analysis system to determine the filter coefficients a_k and the excitation signal. It is considered that representing the vocal tract by the linear prediction (LP) parameters is a sufficient representation and these parameters still produce a high quality speech. This is true for both voiced and unvoiced frames of the speech. For voiced frames, we used a 10th order linear predictive coding model, while for unvoiced frames, we use only a 8th order model, because the speech spectrum of unvoiced sounds is described sufficiently well by the lower order model. The filter essentially represents the vocal tract. For speech sampled at 8 kHz, the determination of the filter coefficients and excitation will usually be every 30 ms (240 samples). In addition to that, it is known that the perfect excitation for the all pole synthesis filter is the residual signal, and a class of linear predictive coders depends on encoding the residual signal efficiently. That means, all the information that has not been captured by linear prediction analysis are carried by the residual excitation such as the phase, pitch information, and zeros due to nasal sounds.

In our proposed method, the encoding of the residual is based on selecting the most important pulses in the residual signal rather than encoding all pulses (samples) by dividing the frame to subframes and making proper classification.

The block diagram of our new model (Encoder) is shown in Figure 3. In this model, we divide the original speech signal in time domain to main frames. After that, we classify these main frames to voiced and unvoiced frames using energy threshold where the energy of voiced frames is higher than this threshold value while the energy of unvoiced frames is below this threshold. If the main frame is voiced, this frame will be filtered to get the residual signal that is also divided to 6 subframes. After that , we make energy

classification to the subframes so that the subframe with higher energy will get more pulses than the subframe with lower energy. If the main frame is unvoiced, the same procedure will be done but there will be no energy classification and all the subframes will have the same number of pulses which is the number chosen for the lowest energy subframe in the voiced frame.
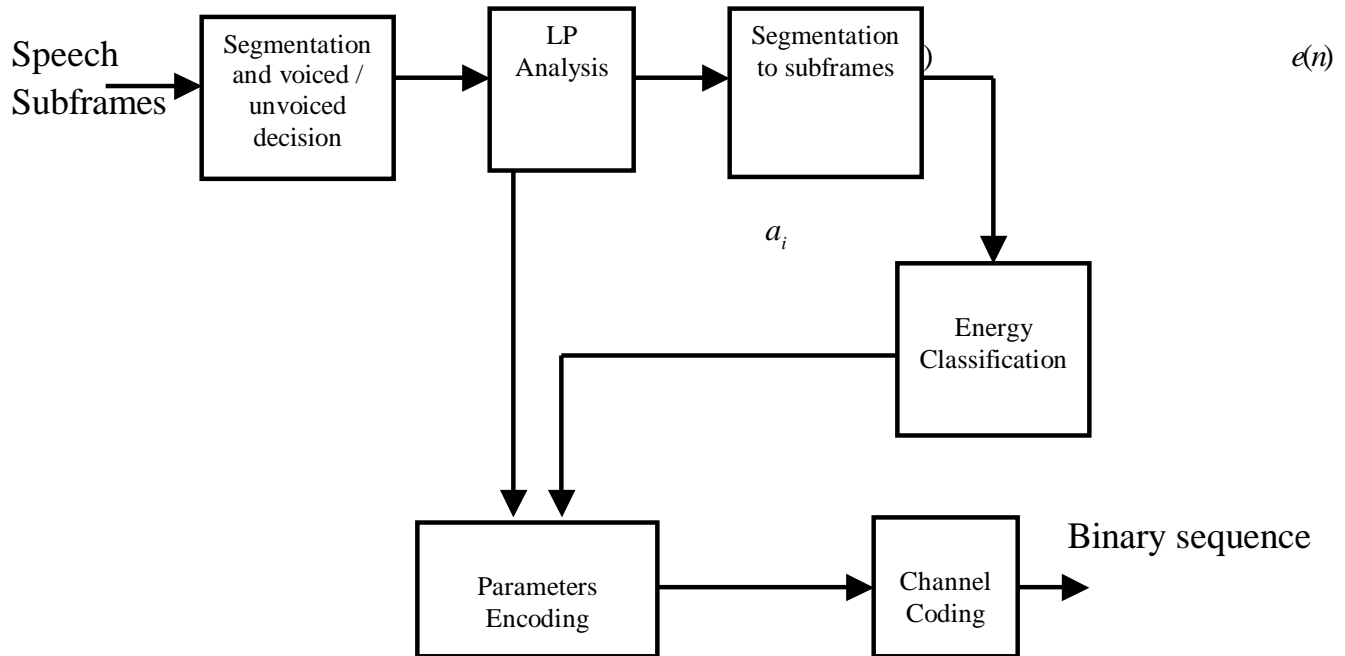
Speech
Subframes

Segmentation and voiced / unvoiced decision

LP Analysis

Segmentation to subframes

$e(n)$

a_i

Energy Classification

Parameters Encoding

Channel Coding

Binary sequence

Figure 3: Encoder

6. The Decoder Stage:

At the decoder, an all pole filter for the synthesis system uses the coefficients obtained from the analysis system and the excitation signal to reconstruct the speech signal as shown in Figure 4. The first step is using voiced / unvoiced information bit. If the frame is voiced, a large number of residual samples are used to reconstruct the residual signal. Then we reconstruct the voiced speech frame by exciting the 10th order filter that uses LPC coefficients by the reconstructed residual signal. If the frame is unvoiced, a less number of residual samples are used to reconstruct the residual signal. Then we reconstruct the unvoiced speech frame by exciting the 8th order filter that uses LPC coefficients by the reconstructed residual signal. Every reconstructed speech frame will be multiplied by the gain which will make the energy of the reconstructed frame to be equal to the energy of the original frame. Then we multiply every frame by a Hanning window and make an overlap between every two consequent frames and add these frames to reconstruct the whole speech signal.
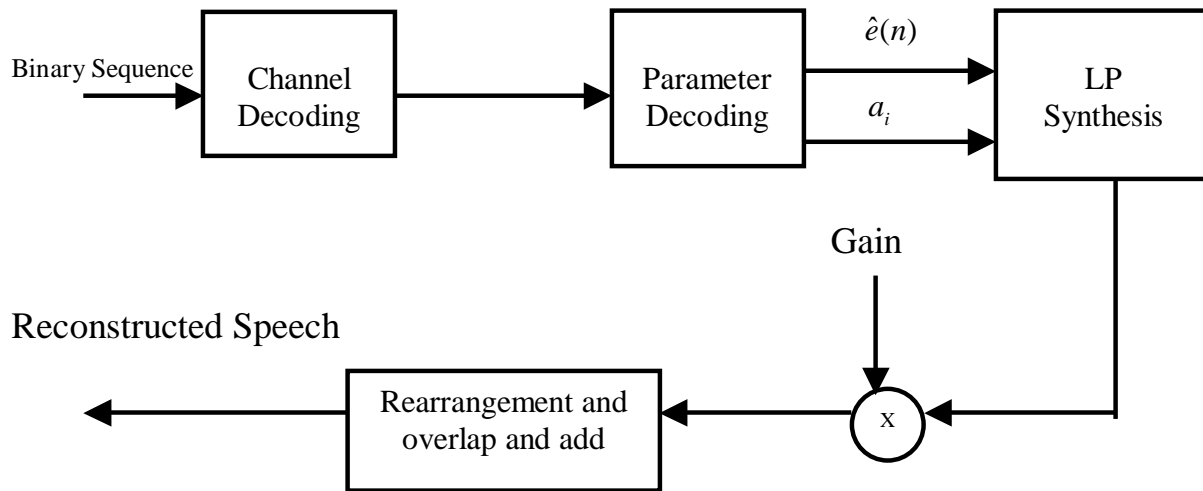
Figure 4: Decoder

In our model, an efficient coder operating between 7 and 9 kbps is presented. The value of the bit rate depends on the energy threshold value. At bit rates above 7 kbps, the speech quality of this coder is high. The high quality comes from the emphasis on coding of the perceptually important residual components.

In the next section, the quantization process will be described for the amplitudes of the pulses and the Line Spectrum Pairs (LSPs).

7. The Quantization of the Residual Pulses:

We used an adapted procedure to encode the amplitudes of the residual signal. This procedure decreases the quantization error and produces a high quality reconstructed speech signal. The key of the algorithm used to encode the sample amplitudes is to find the proper scale to make the length of the amplitudes within the range [1-10] as shown in Table 1, and then representing the amplitudes using 4 bits (the last bit is the sign bit).after scaling to explore the full resolution. This method can be summarized as follows:

Suppose we have amplitudes: a_n = [amp1, amp2,…,ampN] where N is the number of the considered pulses, then:
1. Find the maximum absolute value of the amplitudes in the frame (a_{max}).
2. If a_{max} multiplied by 10^2 is greater than 1, then the first 2 bits in the frame are [00].
3. Else if a_{max} multiplied by 10^3 is greater than 1 and a_{max} multiplied by 10^2 is less than 1 then the first 2 bits in the frame are {01}.
4. Else if a_{max} multiplied by 10^4 is greater than 1 and a_{max} multiplied by 10^3 is less than 1, then the first 2 bits in the frame are {10}.

5. Otherwise, the first 2 bits in the frame are {11}.
6. If the first two bits are {00}, then multiply all the amplitudes in the frame (a_n) by 100, else if the first two bits are {01}, then multiply all the amplitudes in the frame (a_n) by 1000, else if the first two bits are {10}, then multiply all the amplitudes in the frame (a_n) by 10^4, otherwise, multiply all the amplitudes in the frame (a_n) by 10^5 as shown in Table 1.
7. Round the amplitudes in the frame and convert into 4 bits binary number.

Table 1: Scale Code

Scale Code	Range before scaling (Linear)	Range before scaling (Log_{10})	Scale number	Range after scaling and comment
--	$0\text{-}10^{-5}$	$[-\infty,-5]$	--	Ignored
11	$10^{-5}\text{-}10^{-4}$	$[-5,-4]$	10^5	1-10
10	$10^{-4}\text{-}10^{-3}$	$[-4,-3]$	10^4	1-10
01	$10^{-3}\text{-}10^{-2}$	$[-3,-2]$	10^3	1-10
00	$10^{-2}\text{-}10^{-1}$	$[-2,-1]$	10^2	1-10
--	$10^{-1}\text{-}\infty$	$[-1, \infty]$	--	Rare

8. The Quantization of LPC Coefficients using LSFs:

We utilized the procedure described in to encode the LPC coefficients. This procedure decreases the quantization error and produces a high quality reconstructed speech signal.

The Line Spectral Frequencies (LSF) have a well behaved dynamic range. On the other hand, The LSP coefficients represent the LPC model in the frequency domain and lend themselves to a robust and efficient quantization of the LPC parameters.

Therefore, the first LSP represents low frequency components and the last LSP represents high frequency components. Hence, if the LP coefficients are encoded as LSFs, we do not need to spend the same number of bits for each LSF. This is because higher LSFs correspond to the high frequency components and high frequency components have less effect in speech perception.

Therefore, higher LSFs can be quantized using fewer bits than lower LSFs. This reduces the bit rate while keeping the speech quality almost the same.

The LSPs are related to the poles of the LPC filter H(z) (the zeros of the inverse filter A(z)). Usually, the LSFs are more concentrated around formants. Moreover, spectral sensitivity of each LSF is localized. Hence, in order to allow for efficient and robust quantization the LPC parameters are encoded as LSPs. In addition to these, the LSP

parameters are an ordered set of values between 0 and π (ascending order). In this section, we developed a procedure to encode the LSPs coefficients in a robust way with regard to quantization. This procedure can be summarized as follows:
1. Multiply all LSPs by (0.5/ π), to make the range of LSP between 0-0.5.
2. Scale the first LSP to fit into a 6 bits binary number.
3. Convert the result to binary number using (6 bits).
4. Subtract from all other LSPs in the same frame the value found in the step 2 above.
5. Round the result found in the previous step.
6. Convert the value found in the previous step into binary number.
7. Subtract from all other LSPs in the same frame the value found in step 5 above.
8. Repeat steps (5-7) until finish all LSPs.
The number of bits used from LSP(2) to LSP(10) are {6/6/6/5/5/4/3/3/3}.

9. The Total Bit Rate:

We divide the speech signal to main frames with 30 ms duration (240 samples). Then each of these main frames is filtered to get the residual signal that is also divided to 6 subframes. So each subframe has a duration of 5 ms (40 samples).The filtering process is done using 10 LPC coefficients for the voiced frame and using 8 LPC coefficients for the unvoiced frame.
If the main frame is voiced, we make energy classification to the subframes by sorting the energies of these subframes, where the subframe with higher energy get more pulses than the subframe with lower energy. We give each of the highest 2 subframes with respect to energy 13 pulses which means 1/3 of the 40 pulses (pulse rate 1:3).
The selection of the 13 pulses is discussed in the following paragraph:
We divide each of the residual signal in these 2 subframes into 3 residual signals where the 1st residual signal contains the pulses (1,4,7,10,….etc) and the 2nd residual signal contains the pulses (2,5,8,11,…etc) and the 3rd residual signal contains the pulses (3,6,9,12,…etc). After that, we select the residual signal that has the highest energy. If the highest energy is the energy of the 1st residual signal, then the lag and lead bits are {00}. If the highest energy is the energy of the 2nd residual signal, then the lag and lead bits are {01}. If the highest energy is the energy of the 3rd residual signal, then the lag and lead bits are {10}.
We give the next highest subframe with respect to energy 10 pulses which means 1/4 of the 40 pulses (pulse rate 1:4). These 10 pulses are selected in the same procedure for selecting the 13 pulses discussed before while in this case the residual signal in this subframe is divided to 4 residual signals and then we select the highest residual signal with respect to energy and here we also need only two bits for the lag and lead. The next highest subframe is given 8 pulses which means 1/5 of the 40 pulses (pulse rate 1:5). These 8 pulses are also selected in the same procedure for selecting the 13 pulses

discussed before but in this case we need 3 bits for the lag and lead because the residual signal in this subframe is divided into five residual signals. Finally, the lowest 2 subframes get 6 pulses which means 1/6 of the 40 pulses (pulse rate 1:6) and in the same manner we select the best 6 pulses in each of these subframes and 3 bits is needed for the lag and lead for each subframe because the residual signal in each subframe is divided into six residual signals.

Consequently, each of the voiced frames will have 56 pulses and 15 bits for the lag and lead as shown in Table 2. If the main frame is unvoiced, all the subframes will get 6 pulses which mean 1/6 of the 40 pulses (pulse rate 1:6). In this case, we select the best 6 pulses in each of these subframes in the same manner as discussed before and 3 bits is needed for the lag and lead for each subframe because the residual signal in each subframe is divided into six residual signals.

And thus each of the unvoiced frames will have 36 pulses and 18 bits for the lag and lead as shown in Table 3. Finally in Table 4, we show the total bits for each main voiced and unvoiced frame. We use 4 bits for quantizing each pulse and two bits for the scale code. For the case of using 10 LPC coefficients for the voiced main frame, we will need {6/6/6/6/5/5/4/3/3/3} bits for LSP(1) to LSP(10) which means we need 47 bits for representing these coefficients. While in the case for using 8 LPC coefficients for the unvoiced frame, we will need {6/6/5/5/4/4/3/3} bits for LSP(1) to LSP(8) which means we need 36 bits for representing these coefficients. We give 6 bits for energy value which is the energy of the original voiced frame that will be used for the gain at the decoder.

Table 2: Distribution of the pulses and lag and lead bits of the voiced frame

Energy Of Subframe	Pulse Rate	Number of Subframes	Number of Pulses for each Subframe	Subtotal	Number of bits for lag and lead for each Subframe	Number of bits for lag and lead for each row
Highest	1:3	2	13	2*13=26	2	2*2=4
Second Highest	1:4	1	10	1*10=10	2	2*1=2
Third Highest	1:5	1	8	1*8=8	3	3*1=3
Lowest	1:6	2	6	2*6=12	3	3*2=6
Total		6 subframes		56 pulses		15 bits

Table 3: Distribution of the pulses and Lag and lead bits of the unvoiced frame

Energy Of Subframe	Pulse Rate	Number of Subframes	Number of Pulses for each Subframe	Subtotal	Number of bits for lag and lead for each Subframe	Total number of bits for Lag and lead
-	1:6	6 subframes	6 pulses	6*6=36 pulses	3 bits	3*6=18 bits

Table 4: Total Bits for each main voiced and unvoiced frame

	Voiced frame	Unvoiced frame
LP coefficient	47	36
Pulses and scale code	56*4+2=226	36*4+2=146
Lag and lead	15	18
Energy	6	6
Total	294	206

If all frames are voiced frames which is the worst case (this case is impossible), the number of bits for each main frame will equal to 294 bits/frame, and then the data rate (R) can be computed as: R = 294 (bits/frame) * 33.3 (frame/sec) = 9800 bits/sec. Hence, in the worst case, we have a total bit rate equal to 9.8 kbps with high quality reconstructed speech. But with the case of voiced and unvoiced frames, the data rate will be less than 9 kbps.

10. *Advantages of the Model:*

The advantages of our model can be summarized as follows:
- The coder operates between 7 and 9 kbps.
- A good quality is achieved for the reconstructed speech signal.
- The transmitted power is reduced as a result from reducing the bit rate
- There is no dependency on the pitch in our model.
- The model is characterized by its low complexity and small delay.
- It allows for error detection and correction procedure.
- It has an efficient coding procedure.

11. Conclusions:

In this research, we developed a technique to reduce the bit rate while preserving the quality of the speech by extracting some parameters from time domain (LP coefficients, residual pulses, energy value and lag and lead bits).

After that, we used efficient techniques to quantize these parameters, by using two different quantization procedures. Hence, the development of a low bit rate speech coder based on the linear prediction was presented. Computationally efficient techniques and algorithms were presented which have resulted in reconstructed speech of good quality and intelligibility at bit rate from 7 to 9 kbps.

Experimental results proved that this system is graded "good" at its worst level. Hence, in this Model, we presented an efficient vocoder operating between (7 to 9 kbps) depending on energy threshold in the decision process (weather the speech frame is voiced or unvoiced), since this threshold value determines the number of pulses used to encode the residual signal. The speech quality of this vocoder at bit rates above 7 kbps is high because of the emphasis on coding of the perceptually important residual components.

References:

[1] Kang Sangwon, Shin Yongwon, and Fischer Thomas. (2004). "Low-Complexity Predictive Trellis-Coded Quantization of Speech Line Spectral Frequencies". IEEE Transactions on Signal Processing, Vol. 52, No. 7.

[2] Alku Paavo, and Bäckström Tom. (2004). "Linear Predictive Method for Improved Spectral Modeling of Lower Frequencies of Speech With Small Prediction Orders". IEEE Transactions on Speech and Audio Processing, Vol. 12, No. 2.

[3] Atal Bishnu. (1982). "Predictive coding of speech at low bit rates". IEEE Transactions on Communications, COM-30(4):600-614.

[4] Barnwell Thomas. (1980). "Windowless Techniques for LPC Analysis". IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP- 28(4), p. 421.

[5] Brinker Albertus C. den, Voitishchuk, V., and Eijndhoven Stephanus J. L. van. (2004). "IIR-Based Pure Linear Prediction". IEEE Transactions on Speech and Audio Processing, Vol. 12, No. 1.

[6] Viswanathan, V., Higgins, A. and Russell, W. (1982). "Design of a Robust Baseband LPC Coder of Speech Transmission Over 9.6 Kbit/s Noisy Channels". IEEE Transactions on Communications, Vol. COM- 30(4), p. 663.

[7] Chong Nicola, and Cox Richard. (2003). "An Intelligibility Enhancement for the Mixed Excitation Linear Prediction Speech Coder". IEEE Signal Processing Letters, Vol. 10, No. 9.

[8] Galand, C., Lancon, E., Rosso, M. and Menez, J. (1986). "A New Architecture of Multi-pulse Excited Linear Predictive Coder". Elsevier Science Publishers.

[9] Gray, A. and Wong, D. (1980). "The Burg Algorithm for LPC Speech Analysis/Synthesis". IEEE Transactions on Acoustics, Speech and Signal Processing, VOL. ASSP- 28(6), p. 609.

[10] Härmä Aki. (2001). "Linear Predictive Coding With Modified Filter Structures". IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 8.

[11] Hu Hwai-Tsu, and Wu Hsi-Tsung. (2000). "A Glottal-Excited Linear Prediction (GELP) Model for Low-Bit-Rate Speech Coding", Proc. Natl. Sci, Counc. ROC(A) Vol. 24. pp. 134-142.

[12] Johansson Anders. (1986). "Linear Prediction of Speech Using Recursive Analysis", Elsevier Science Publishers.

[13] Proakis John and Manolakis D. (1996). "Digital Signal Processing: Principles, Algorithms and Applications" , third edition, Macmillan, New York.

[14] Spanias Andreas. (1994). "Speech coding: a tutorial review". Proc. IEEE, Vol. 82, pp. 1541-1582.

[15] Papamichalis Panos. (1987). "Practical Approaches to Speech Coding", Prentice Hall, Inc. Texas Instruments, Inc. Rice University.

[16] Proakis John. (2001). "Digital Communications". Fourth Edition. McGraw-Hill, New York.

[17] Ingle Vinay, and Proakis John. (2000). "Digital Signal Processing Using MATLAB". BookWare Companion Series, Northeastern University.

[18] Hu Jianling, Xu Sheng, and Chen Jian. (2001). "A Modified Pitch Detection Algorithm". IEEE Communications Letters, Vol. 5, No. 2.

[19] CCITT Recommendation G.721. (1984). "32kb/s Adaptive Differential Pulse Code Modulation (ADPCM)". CCITT.

[20] Kim Hong Kook, and Lee Hwang Soo. (1999). "Interlacing Properties of Line Spectrum Pair Frequencies". IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 1.

[21] Kim Hong Kook, Choi Seung Ho, and Lee Hwang Soo. (2000). "On Approximating Line Spectral Frequencies to LPC Cepstral Coefficients". IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 2.