# Speech Signal Compression Analyses Based Prediction And Coding

[1]Khalid .A. Al Smadi, [2]Farouk M Al-Taweel.
[1]Jordanian Sudanese College for Science & Technology., Sudan,
[2]Department of Electrical Engineering, Al-Balqa Applied University, Jordan

Article Info

Received: 5/12/2011
Accepted:10/1/ 2012
Published online: 1/6/2013

ISSN 2231-8844

**Abstract**
In this paper, investigated to compute the effect of the Digital compression of speech signal based prediction and coding, can increase the speed of data transmission using coding one type of such coding is Differential pulse-code modulation in the latter case. The present study provides linear predicting analysis consists in that that speech; the sample can be approximate as a linear combination of last samples. Linear predictor the model provides a sensible, reliable and exact method in order that it estimated parameters it Characterize linear, time variable system. In this project we carry out an uneasy voice LPC voice coder for low speech compression Conclusion: in sufficient detail the method of digital compression of the speech signal based on linear prediction. It is shown that there are several approaches to solving this problem

***Keyword: Speech signal, coding, compression, modulation***.

## 1. Introduction

Speech is a very basic way for humans to convey information to one another. With a bandwidth of only4kHz, speech can convey information with the emotion of a human voice. People want to be able to hear someone's voice from anywhere in the world-as if the person would be in the same room. Speech can be defined such as the response of the vocal tract to one or more excitation signals (Gibson, 2006).

Predicted by the current value of the reference on the basis of previous M samples, for concreteness, assume that means the current reading the source, and let denote the predicted value (assessment) for which is defined a

$$\hat{x}_n = \sum_{k=1}^{M} a_k \cdot x_{n-k} \tag{1}$$

We illustrate the above on segmented speech signal see Fig.1

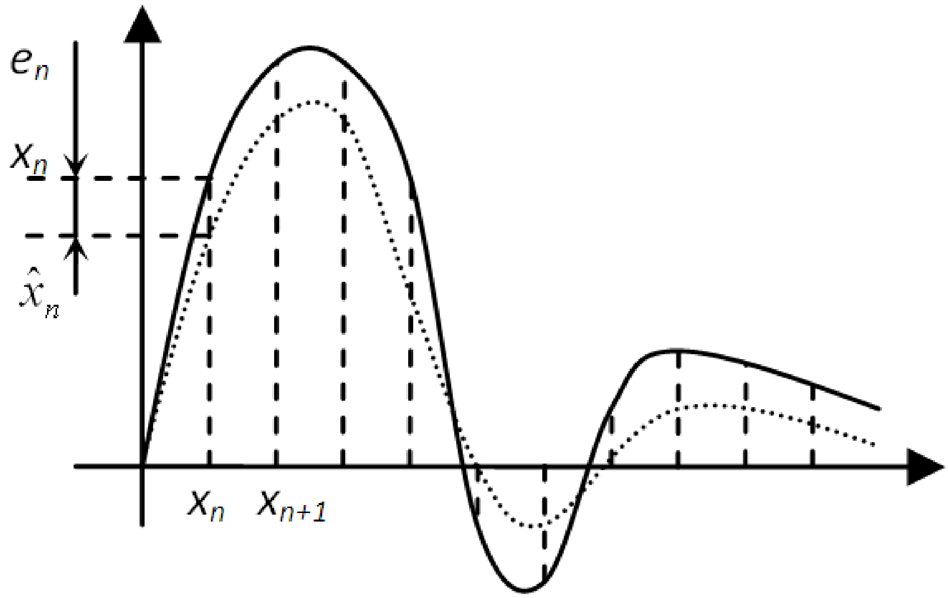



Fig.1: segmented speech signal

That is a linear function of previous samples, with "nonlinear" prediction - this is a nonlinear function. The order of prediction is determined by the number of previous samples used. That is, the prediction of zero and first order is linear, and the second and higher order - non-linear. In the linear prediction to restore the signal is much easier than with a nonlinear prediction. We will consider only the linear prediction (Yu et al., 2006).

## 2. Material and methods

a. The prediction of zero order In this case, to predict the current frame using only previous count Fig.2.

$$\hat{x}_n = x_{n-1} \Rightarrow e_n = x_n - \hat{x}_n = x_n - x_{n-1} \tag{2}$$

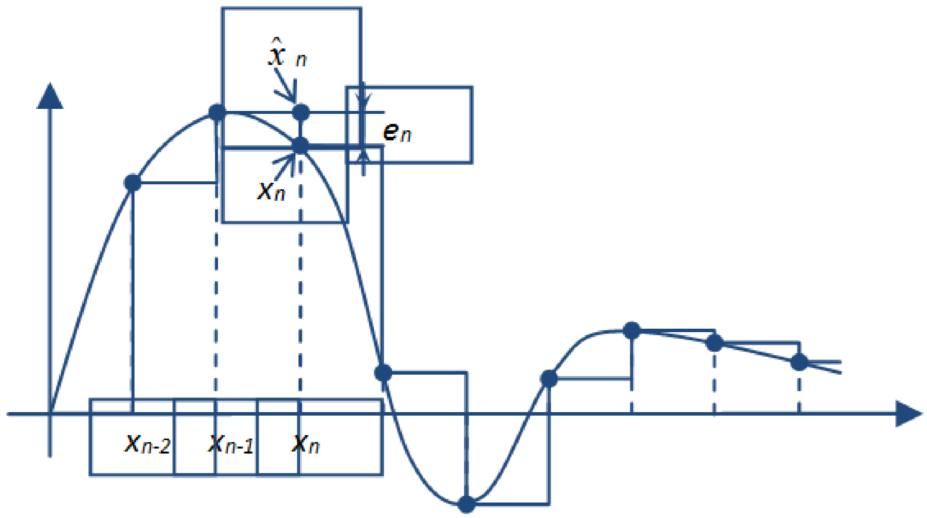



Fig.2: Prediction of zero order

(b) Prediction of the first order (linear extrapolation) to predict the current frame is not only used the previous count, but the difference between the penultimate and last counts, which is added to the total result in equation (3) as shown in Fig.3

$$\hat{x}_n = x_{n-1} + \Delta x = x_{n-1} + (x_{n-1} - x_{n-2}) = 2 \cdot x_{n-1} - x_{n-2}$$

$$e_n = x_n - \hat{x}_n = x_n - 2 \cdot x_{n-1} + x_{n-2}. \tag{3}$$

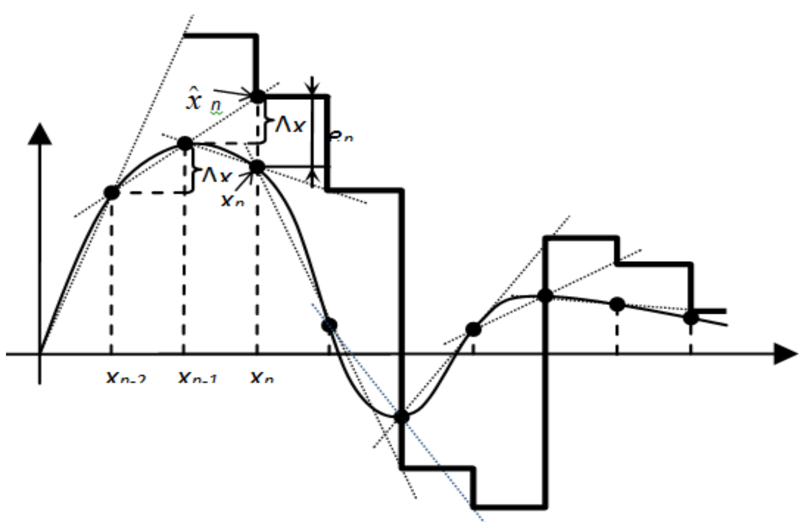



Fig.3: Prediction of the first order

### 2.1 The coefficients of linear prediction

Signal error when using a linear prediction is equivalent to the passage of the original signal through a linear digital filter. This filter is called a filter error signal or the inverse filter. Denote the transfer function of this filter as a (z) as show in: equation (4)

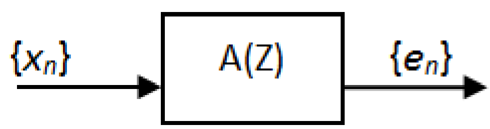



Fig.4: Inverse filter

$$A(z) = \frac{E(z)}{X(z)} \tag{4}$$

Where E (z) and X (z) - a direct z - transform of the error signal and input signal, respectively.

At the receiving end of an error signal when passing through the shaping filter (FF), we should ideally obtain the original signal. Denote the transfer function as a shaping filter K (z). Ie transfer function K (z) associated with A (z) the following relation (Ekman et al., 2008).

$$K(z) = \frac{1}{A(z)} = \frac{X(z)}{E(z)} \tag{5}$$

Let us consider the combined encoder and decoder as shown in Figure 5

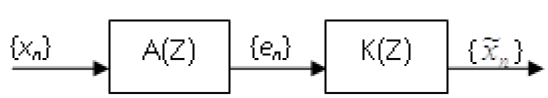



Fig.5: combined encoder and decoder

Provided that A (z) K (z) = 1, will be provided absolutely accurate signal reconstruction. But this is the ideal; in fact, this cannot be, for reasons which are discussed below. For example, we find the transfer functions of the FSO and FF for different types of linear

prediction.
a) The prediction of zero order; as in Equations (6) and (7);

$$A(z) = \frac{E(z)}{X(z)} = \frac{X(z) - z^{-1} \cdot X(z)}{X(z)} = 1 - z^{-1}. \tag{6}$$

$$K(z) = \frac{1}{A(z)} = \frac{1}{1 - z^{-1}} = \frac{(z-0)}{(z-1)} \tag{7}$$

We have shown that the filter is unstable (stability limit), since the pole is located on the unit circle.
b) Prediction of the first order;

$$A(z) = \frac{E(z)}{X(z)} = \frac{X(z) - 2 \cdot z^{-1} \cdot X(z) + z^{-2} \cdot X(z)}{X(z)} = 1 - 2 \cdot z^{-1} + z^{-2} \tag{8}$$

$$K(z) = \frac{1}{1 - 2 \cdot z^{-1} + z^{-2}} = \frac{(z-0)^2}{(z-1)^2} \tag{9}$$

We have shown that such a filter is also unstable (stability limit).

c) The general form of prediction as in equations (10) and (11);

$$\hat{x}_n = \sum_{k=1}^{M} a_k \cdot x_{n-k}$$

$$e_n = x_n - \hat{x}_n = x_n - \sum_{k=1}^{M} a_k \cdot x_{n-k} \tag{10}$$

$$A(z) = \frac{E(z)}{X(z)} = \frac{X(z) - \sum_{k=1}^{M} a_k \cdot z^{-k} \cdot X(z)}{X(z)} = 1 - \sum_{k=1}^{M} a_k \cdot z^{-k}\,; \quad K(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^{M} a_k \cdot z^{-k}} \tag{11}$$

Based on the above examples the following conclusions (Magi et al., 2009). Filter the error signal is always a FIR filter, and the shaping filter - IIR filter. The coefficients of the transfer function FF, which, as already mentioned, are the coefficients of linear prediction (LPC: Linear Prediction Coefficients), should be such that: (a) Shaping filter is stable; (b) Error was minimal.

To obtain the transfer function of FF, the most accurately reproducing the frequency response of the vocal tract for the sound, you should determine the coefficients of the transfer function from the condition of least error of linear prediction of speech signal (Wipf & Nagarajan, 2010). (on the condition of minimum mean square error). Write an expression for estimating the variance of the error signal, which must be minimized:

$$\hat{\sigma}_e^2 = \frac{1}{N} \cdot \sum_{n=1}^{N} (x_n - \hat{x}_n)^2 = \min \tag{12}$$

$$s^2 = \sum_{n=1}^{N} (x_n - \hat{x}_n)^2 = \sum_{n=1}^{N} (x_n - \sum_{k=1}^{M} a_k \cdot x_{n-k})^2 = \min \tag{13}$$

We have shown that $s^2 = f(a_1, a_2, a_3, ..., a_m)$

Differentiate and equate its partial derivatives to find extreme:

$$\frac{\partial s^2}{\partial a_m} = 0, m = \overline{1, M}. \tag{14}$$

$$\frac{\partial s^2}{\partial a_m} = \sum_n 2 \cdot (x_n - \sum_{k=1}^{M} a_k \cdot x_{n-k})^2 \cdot \left( - \sum_{k=1}^{M} \frac{\partial a_k}{\partial a_m} \cdot x_{n-k} \right) = 0 \tag{15}$$

$$\frac{\partial a_k}{\partial a_m} = \begin{cases} 1, k = m \\ 0, k \neq m \end{cases} = \delta_{km} \sum_{k=1}^{M} \delta_{km} \cdot x_{n-k} = x_{n-m} \tag{16}$$

$$\frac{\partial s^2}{\partial a_m} = \sum_n 2 \cdot (x_n - \sum_{k=1}^{M} a_k \cdot x_{n-k}) \cdot (- x_{n-m}) = 0$$

Then,

$$\frac{\partial s^2}{\partial a_m} = \sum_n (x_n - \sum_{k=1}^{M} a_k \cdot x_{n-k}) \cdot (x_{n-m}) = 0 \tag{17}$$

$$\sum_n (x_n - \sum_{k=1}^{M} a_k \cdot x_{n-k}) \cdot (x_{n-m}) =$$

$$\sum_n (x_n \cdot x_{n-m} - \sum_{k=1}^{M} a_k \cdot x_{n-k} \cdot x_{n-m}) =$$

$$\sum_n x_n \cdot x_{n-m} - \sum_n \sum_{k=1}^{M} a_k \cdot x_{n-k} \cdot x_{n-m} = 0 \tag{18}$$

$$\varphi(k,m) = \sum_n x_{n-k} \cdot x_{n-m} = \varphi(0,m) = \sum_{k=1}^{M} a_k \cdot \varphi(k,m \tag{19}$$

To calculate the functions necessary to define the limits of summation over n:, where N - number of samples in a segment of the PC, and M - number of samples required to calculate the coefficients of the prediction (M + 1) the frame. So, the first

$$\hat{x}_n = f(x_{n-1}, x_{n-2}, x_{n-3}, ..., x_{n-M}), \ n = M + 1 \tag{20}$$

$$\varphi(k,m) = \sum_{n=M+1}^{N} x_{n-k} \cdot x_{n-m} \tag{21}$$

$n - k = j$ =>$n = k + j$, $n - m = k + j - m$<=>$n - m = i + j$, $i = k - m$.

$$\varphi(k,m) = \sum_{j=M+1-k}^{N-k} x_j \cdot x_{j+i} \tag{22}$$

Thus, we have an expression that has the structure of short-term normalized, but depends not only on the relative shift of the sequence i, but also on the position of the sequences within the segment, defined the index k, within the limits of summation. This method of determining the function is called covariance.

Expression (*) is a system of linear equations) with respect to all of whose coefficients different.

When using the covariance method obtained unbiased estimates of the coefficients of linear prediction, E{$a_k$}= $a_{k.ucm}$= where $a_{k.ucm}$. - The true values of the coefficients of linear prediction. In this procedure, as intermediate values, using some of the coefficients km, which are called reflection coefficients. Their physical meaning is as follows. Human vocal tract is a tube consisting of sections connected in series, but with different diameters. With the passage of sound waves through a system of reflection occur at the junction of sections, each joint is heterogeneity. The reflection coefficient characterizes the cross-junction of two sections of the reflection coefficient is equal to (Sreenivas et al., 2009)

$$r_m = \frac{s_{m+1} - s_m}{s_{m+1} + s_m}; \quad -1 \le r_m \le 1; \quad \Leftarrow \quad s_i \ge 0 \tag{23}$$

Explain its meaning in Figure 6.


Figure 6

If $r_m$ = -1, then will break in the chain of transmission (open direct branches). This should not be, so you must watch this. Model of acoustic tube can be represented as a filter having a lattice (or ladder) structure. The main parameters of this filter are the reflection coefficients. Since the reflection coefficients and prediction coefficients are calculated within the same procedure, the algorithm of the Levinson-Durbin, they can be expressed in terms of each other. We present here the algorithm. Direct recursion (reflection coefficients of the prediction coefficients.

$$\left[\begin{array}{l} \left.\begin{array}{c} a_m^{(m)} = -r_m \\ a_j^{(m)} = a_j^{(m-1)} + r_m \cdot a_{m-j}^{(m-1)},\ j = \overline{1, m-1} \end{array}\right\}, m = \overline{1, M} \\ \quad a_j = a_j^{(M)},\ j = \overline{1, M} \end{array}\right. \tag{24}$$

Contact recursion (prediction coefficients ◊ reflection coefficients):

$$\left[\begin{array}{l} \quad a_j^{(M)} = a_j,\ j = \overline{1, M} \\ \left.\begin{array}{c} r_m = -a_m^{(m)} \\ a_j^{(m-1)} = \dfrac{a_j^{(m)} + a_m^{(m)} \cdot a_{m-j}^{(m)}}{1 - r_m^2},\ j = \overline{1, m-1} \end{array}\right\}, m = \overline{M, 1}. \end{array}\right. \tag{25}$$

The error signal is a filter or FIR filter, which means that no branch of feedback. System may also have a strictly linear. Linearity is a very important factor with regard to MS in those cases where you want to keep relative positions of the elements of the signal. This greatly simplifies the task of designing and allows you to pay attention only approximations of their response. For this advantage has to pay the necessity of an extended approximation of the impulse response in the case of filters with steep (Sreenivas et al., 2009). We represent the graph filter having a lattice structure, the example of a filter of order 3: In contrast to the shaping filter, this filter has one input and two outputs:
1) $e_i$ - a sequence of sampling error signal direct linear prediction;
2) $b_i$ - a sequence of sampling error signal backward linear prediction.

$$b_{n-1} = x_{n-1} - \hat{x}_{n-1} \tag{26}$$

The importance of $b_i$ in the fact that through it together with the signal error ei can be estimated reflection coefficients.

$$r_m = -\frac{\sum_{n=1}^{N} e_n^{(m-1)} \cdot b_n^{(m-1)}}{\sqrt{\sum_{n=1}^{N} \left(e_n^{(m-1)}\right)^2} \cdot \sqrt{\sum_{n=1}^{N} \left(b_n^{(m-1)}\right)^2}} \tag{27}$$

The resulting formula for calculating the reflection coefficient also has another physical meaning. This is nothing, as the correlation coefficient between the sequence of samples of the error signal the forward and backward linear prediction. We also give recursive difference equations of a lattice filter error signal

$$\left\{\begin{matrix} e_n^{(m)} = e_n^{(m-1)} + r_m \cdot b_{n-1}^{(m-1)} \\ b_n^{(m)} = b_{n-1}^{(m-1)} + r_m \cdot e_n^{(m-1)} \end{matrix}\right\}, m = \overline{1, M} \qquad e_n^{(0)} = x_n ; b_n^{(0)} = x_n \tag{28}$$

3. **Experimental design**

Having a method for determining the coefficients of prediction, consider the block diagram of a practical system, as in Fig.4

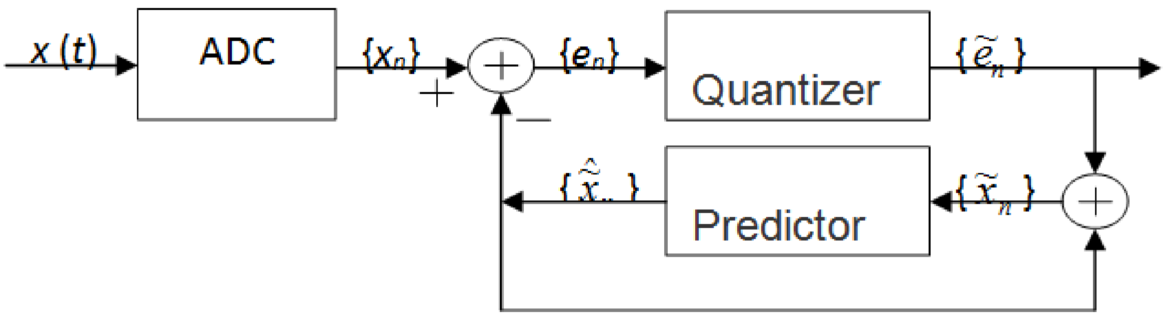



Fig.7

This scheme is a predictor in the feedback loop, covering the Quantizer. Log predictor identified. It is a countdown signal, distorted by the quantization error signal. Output predictor is fig 8.

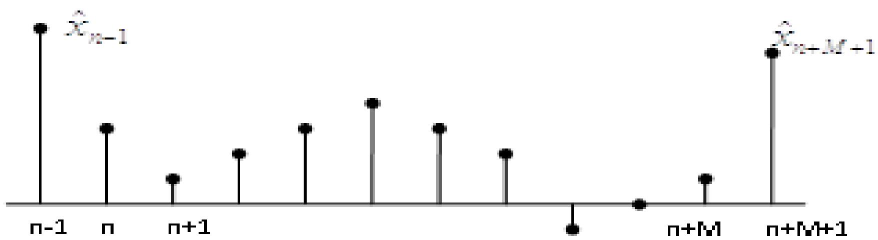



Figure 8

$$\hat{\tilde{x}}_n = \sum_{k=1}^{M} a_k \cdot \tilde{x}_{n-k} \tag{29}$$

At the point of reception using the same predictor as to the transfer, and its output is $\hat{\tilde{x}}_n$ summed s $\tilde{e}_n$ to get $\tilde{x}_n$ (Figure.9)

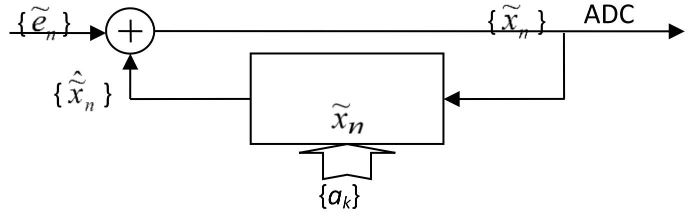



Fig.9

Improving the quality of estimates can be obtained at the inclusion of a linear filtered estimate past values of the quantized error. Specifically, the estimate can be expressed as follows:

$$\hat{\tilde{x}}_n = \sum_{k=1}^{m} a_k \cdot \tilde{x}_{n-k} + \sum_{k=1}^{l} b_k \cdot \tilde{e}_{n-k} \qquad 30$$

3.1 **Adaptive differential pulse-code modulation**

Improvement that reduces the dynamic range of the quantization noise - is the use of an adaptive Quantizer. Other - do adaptive predictor. The coefficients of the predictor may from time to time to reflect the changing statistics of the signal source. for solutions which uses an algorithm Levinson - Durbin, remains true to the short-term evaluation of the autocorrelation function B (i) (with notation B (i) - already set in place of short-term evaluation of the correlation function of the ensemble. Defined in this way can be a predictor coefficients together with the error quantization handed receiver that uses the same predictor. Unfortunately, the transmission coefficients of the predictor increases the required bit rate, partially offsetting the decrease in the rate achieved by the quantizer with a few bits (a few quantization levels) to reduce the dynamic range of errors obtained by the adaptive prediction. As an alternative predictor of the receiver can calculate its own coefficients and prediction through which (Rajesh et al., 2011)

$$\tilde{x}_n = \tilde{e}_n + \sum_{k=1}^{M} a_k \cdot \tilde{x}_{n-k} \qquad (31)$$

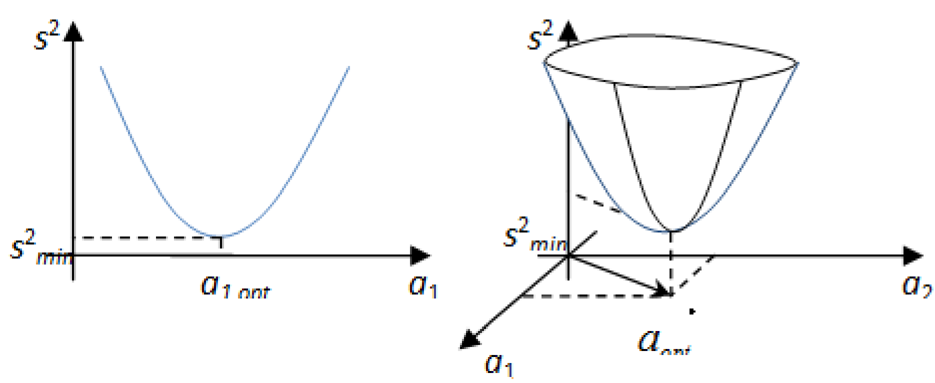



Fig.10 receiver can calculate its own coefficients and prediction

If we neglect the quantization noise is equivalent. Consequently, it can be used to estimate the ACF B (i) in the receiver, and the resulting estimates can be used instead of B (i) in finding the coefficients of the predictor. For a sufficiently large number of quantization levels and the difference between very small. Consequently, the estimate B (i), obtained through, can be used to determine the coefficients of the predictor. Performed so adaptive predictor leads to a low encoding rate of the source data. Instead of using the block processing for finding the coefficients of the predictor {}, as described above, we can adapt the predictor coefficients, using an algorithm of gradient type, which we consider. The main advantage of this method of adaptation - a rejection of the decision, which significantly reduces the computational cost. We write the estimate of the mean squared prediction error:

$$s^2 = \sum_{n=1}^{N}(\tilde{x}_n - \sum_{k=1}^{M} a_k \cdot \tilde{x}_{n-k})^2 = \min \quad (32)$$

Depict two graphs that explain the functional relationship

$$s^2 = f(a_1, a_2, a_3, ..., a_m)$$

$$(s^2 = f(a_1) \text{ , } (s^2 = f(a_1, a_2) \quad (33)$$

Obtained by two factors predict, will become a multi-dimensional parabolic. The purpose of the gradient method is to find a vector aort in which the function s2 will be the smallest value, ie after certain iterations necessary to reach the top of the parabolic. The algorithm of the gradient method is as follows (Hu & Loizou, 2008):

$$\vec{a}(i+1) = \vec{a}(i) - \mu \cdot grad(s^2(i)) \quad (34)$$

$$\vec{a}(i) = \begin{pmatrix} a_1(i) \\ a_2(i) \\ \vdots \\ a_M(i) \end{pmatrix}; \quad grad(s^2(i)) = \begin{pmatrix} \frac{\partial s^2}{\partial a_1} \\ \frac{\partial s^2}{\partial a_2} \\ \vdots \\ \frac{\partial s^2}{\partial a_M} \end{pmatrix} \quad (35)$$

For a small step of the algorithm, we almost completely eliminate the possibility of divergence of the algorithm, but lose in the convergence rate or the rate of finding the coefficients of the predictor. And vice versa. This algorithm converges for a very large number of iterations, in general, when the number of iterations tends to infinity. It is therefore important to ask before computing an allowance, which we can arrange that. Thus, the partial derivative:

$$\frac{\partial s^2}{\partial a_m} = -2\cdot\sum_n(\tilde{x}_n - \sum_{k=1}^{M} a_k(i)\cdot\tilde{x}_{n-k})\cdot\tilde{x}_{n-m} = -2\cdot\left[\sum_n \tilde{x}_n \cdot\tilde{x}_{n-m} - \sum_{k=1}^{M} a_k(i)\cdot\tilde{x}_{n-k}\cdot\tilde{x}_{n-m}\right] =$$
$$= -2\cdot\left[\varphi(0,m) - \sum_{k=1}^{M} a_k(i)\cdot\varphi(k,m)\right] \quad (36)$$

Then the algorithm to adapt the coefficients of linear prediction takes the following form:

| | |
|---|---|
| $a_m(i+1) = a_m(i) - \mu\cdot\frac{\partial s^2(i)}{\partial a_m} = a_m(i) + 2\mu\cdot\left[\varphi(0,m) - \sum_{k=1}^{M} a_k(i)\cdot\varphi(k,m)\right], \quad m = \overline{1,M}$ | (37) |

The following are illustrations of one of the experiments carried Treated segment of the speech signal see Fig.11a and the Prediction error in Fig.11b.

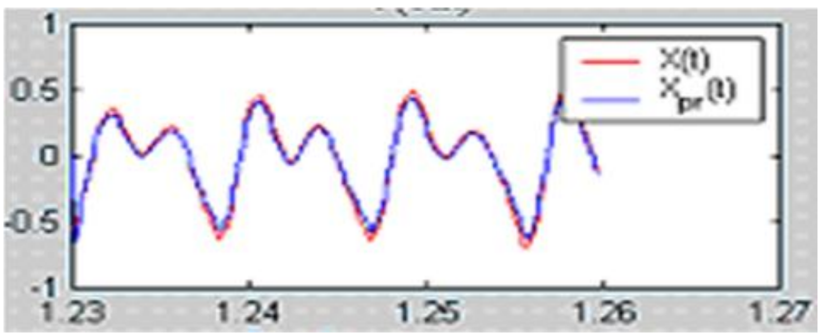



Fig.11a: Treated segment of the speech signal

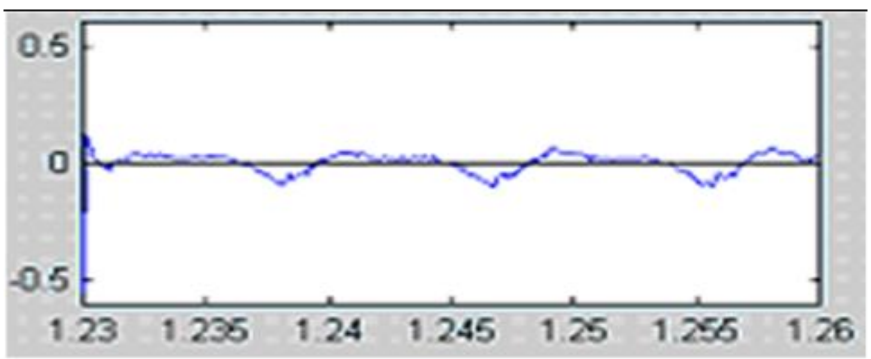



Fig.11b: Prediction error

The reflection coefficients and impulse response of the shaping filter and The resultant segment speech signal shown in Fig 12a. Fig.12.b

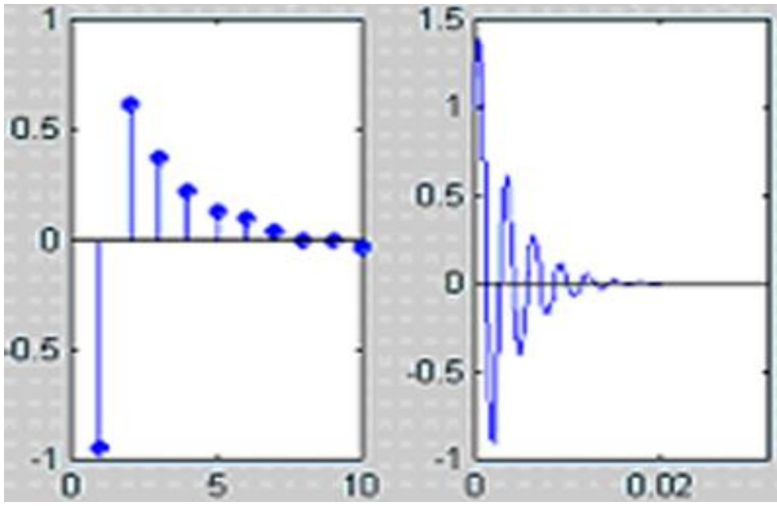
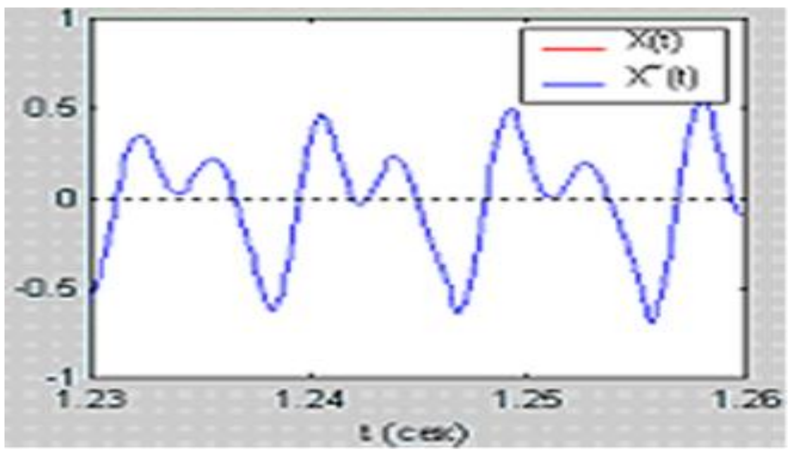



Fig 12a: The reflection coefficients and impulse response of the shaping filter




(b)

Fig.12.b: The resultant segment speech signal

## 4. Results and discussions

Lattice filter is always stable and reflection coefficients are always less than 1, because the reflection coefficients are also the correlation coefficients. The stability of a lattice filter is normal to the bit of reflection coefficients. Bit reflection coefficient affects only the form of the transfer function and, consequently, the diagram of the poles and the impulse response, and the shape of the synthesized MS affects only very slightly, at a constant, relatively high (12) bit error signal (Resch et al., 2007).

In the case of a fixed, relatively low, the bit of reflection coefficients (4) and decreasing the bit error signal to a value of (6), degradation of the synthesized PC is negligible. If the number of bits is less than (6) is already beginning to be significant distortion. If we compare this experience with that, done over a standard filter, then the same segment and with a value of bits (8), there was instability of the synthesized filter and as a consequence, a complete distortion of the speech signal, If the two filters were stable and their capacity factors, as well as bit error signal was the same, then the synthesized signal is identical.

It should also be noted not only the effect of word length ratios predict / reflect on the synthesized signal, but, above all, the very realization of the original analog speech signal , as the basis on which the coefficients themselves are calculated. Therefore, you must have a supply of the bit prediction coefficients to a standard filter for some implementations are not proved to be unstable ,Experimentally, was picked up the option of choosing the bit prediction coefficients (12), and the error signal (8) (bit of reflection coefficients play virtually no role). It is well distinguishable speech.

## Conclusion

In this work, in sufficient detail the method of digital compression of the speech signal based on linear prediction. It is shown that there are several approaches to solving this problem. We give an illustration of the laboratory work done with all the necessary comments and conclusions.

## Reference

Gibson, J. D. (2005). Speech coding methods, standards, and applications. Circuits and Systems Magazine, IEEE, 5(4), 30-49.

Yu, J., Amores, J., Sebe, N., & Tian, Q. (2006, July). A new study on distance metrics as similarity measurement. In Multimedia and Expo, 2006 IEEE International Conference on (pp. 533-536). IEEE.

Ekman, L. A., Kleijn, W. B., & Murthi, M. N. (2008). Regularized linear prediction of speech. Audio, Speech, and Language Processing, IEEE Transactions on, 16(1), 65-73.

Magi, C., Pohjalainen, J., Bäckström, T., & Alku, P. (2009). Stabilised weighted linear prediction. Speech Communication, 51(5), 401-411. James D. Broesch, Dag Stranneby and William Walker. Digital Signal Processing: Instant access. Butterworth-Heinemann. p. 3.

Wipf, D., & Nagarajan, S. (2010). Iterative reweighted l1 and l2 methods for finding sparse solutions. IEEE Journal of Selected Topics in Signal Processing, 4(2), 317-329.

Sreenivas, T. Sreenivas, T. V., & Kleijn, W. B. (2009, April). Compressive sensing for sparsely excited speech signals. In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on (pp. 4125-4128). IEEE

Rajesh, G., Kumar, A., & Ranjeet, K. (2011, September). Speech compression using different transform techniques. In Computer and Communication Technology (ICCCT), 2011 2nd International Conference on (pp. 146-151). IEEE.

Hu, Y., & Loizou, P. C. (2008). Evaluation of objective quality measures for speech enhancement. *Audio, Speech, and Language Processing, IEEE Transactions on*, *16*(1), 229-238.

Resch, B., Nilsson, M., Ekman, A., & Kleijn, W. B. (2007). Estimation of the instantaneous pitch of speech. Audio, Speech, and Language Processing, IEEE Transactions on, 15(3), 813-822.