Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 53(67), Fascicola 2, 2008

An Important Property of the Time-Domain Interpretation for the LSF Parameters

Septimiu Mischie[1], Gabriel Vasiu[1]

Abstract – This paper presents an important improvement that can be obtained by using two new types of linear prediction. These were previously introduced in order to find a time domain interpretation to LSF parameters. We show that the minimum error energy of at least one of the two predictors is much lower than the error of conventional linear prediction.
Keywords: Linear predictive coding (LPC), minimum error energy, low-pass filter, high-pass filter.

I. INTRODUCTION

Linear predictive coding (LPC) [1] is a well-known technique used to speech coding. By using this technique, speech signal is decomposed in LPC coefficients and residual (error) signal. These two components are separate quantized. LPC coefficients are transformed in Line Spectrum Frequency (LSF) parameters[2], that can be better quantized. In [2] is presented a time-domain interpretation of LSF parameters. Namely, these parameters can be obtained using two new types of linear predictors: one is preceded by a low-pass filtering of the input signal and the other is preceded by a high-pass filtering. In this paper, first the expressions of the minimum error energy for two types of predictions are demonstrated. Then, it is demonstrated that both the minimum errors depend on spectral components of the analyzed signal. Furthermore, at least one of the two errors is much lower than the error in conventional LPC predictor.
The paper is organized as follows. In section II the two new types of predictors are presented. In section III the expressions of the minimum error energy are demonstrated and in section IV experimental results are presented.

II. LINEAR PREDICTION USING AVERAGED AND DIFERENTIATED VALUES

The conventional linear prediction of p order considers that a sample of a sequence $x(n)$ can be estimated by a linear combination of the previous p samples,

$$\tilde{x}(n) = -\sum_{i=1}^{p} a_i x(n-i) \qquad (1)$$

where a_i represents LPC coefficients. The prediction error sequence represents the difference between the input sequence and the estimated sequence.

$$e(n) = x(n) - \tilde{x}(n) = x(n) + \sum_{i=1}^{p} a_i x(n-i), \qquad (2)$$

The LPC coefficients are obtained by minimizing the linear prediction error energy, $\sum_n e^2(n)$. That means solving of the following equations

$$\frac{\partial}{\partial a_j}\left[\sum_n e^2(n)\right] = 0,\ j = 1,\dots,p \qquad (3)$$

One of the most used methods to solve the equations (3) is the autocorrelation method. Thus, the elements of the input sequence $x(n)$ are assumed to be different to 0 for n=0,1,…,N–1, and equal to 0 outside this interval. In this way, the following equations are obtained [1]

$$\sum_{i=1}^{p} a_i R(i-j) = -R(j),\ j = 1,2,\dots p\,. \qquad (4)$$

where $R(i)$ represent elements of the autocorrelation sequence of $x(n)$. Also the minimum error energy can be expressed by [1]

$$E_{p,\min} = R(0) + \sum_{i=1}^{p} a_i R(i) \qquad (5)$$

In [2], two new types of linear prediction are presented. First, two sequences are introduced

$$x^+(n) = \frac{1}{2}\left[x(n) + x(n+1)\right] \qquad (6)$$

$$x^-(n) = \frac{1}{2}\left[x(n) - x(n+1)\right] \qquad (7)$$

and then the elements of input sequence are estimated as follows

$$\tilde{x}^+(n) = -\sum_{i=1}^{p} h_i^+ x^+(n-i) \qquad (8)$$

$$\tilde{x}^-(n) = -\sum_{i=1}^{p} h_i^- x^-(n-i)\,. \qquad (9)$$

[1] Facultatea de Electronică şi Telecomunicaţii, Departamentul MEO Bd. V. Pârvan Nr. 2, 300223 Timişoara, e-mail: septimiu.mischie@etc.upt.ro, gabriel.vasiu@etc.upt.ro

Thus, expression (8) represents a linear prediction depending on average value of two consecutive values of $x(n)$ and expression (9) represents a linear prediction depending on differentiated value of two consecutive values of $x(n)$.
It follows the expressions of linear prediction error sequences:

$$e^+(n) = x(n) - \tilde{x}^+(n) = x(n) + \sum_{i=1}^{p} h_i^+ x^+(n-i) = x(n) + \sum_{i=1}^{p} \frac{h_i^+}{2}\left[x(n-i) + x(n-i+1)\right] \quad (10)$$

and, respectively,

$$e^-(n) = x(n) - \tilde{x}^-(n) = x(n) + \sum_{i=1}^{p} h_i^- x^-(n-i) = x(n) + \sum_{i=1}^{p} \frac{h_i^-}{2}\left[x(n-i) - x(n-i+1)\right]. \quad (11)$$

The coefficients h_i^+ and respectively, h_i^- can be determined similar with a_i by minimizing the linear prediction error energy.
In this way the following equations are obtained [2]:

$$\sum_{i=1}^{p} h_i^+ \left[2R(i-j) + R(i-j+1) + R(i-j-1)\right] = -2\left[R(j) + R(j-1)\right], \quad 1 \le j \le p \quad (12)$$

and, respectively,

$$\sum_{i=1}^{p} h_i^- \left[2R(i-j) - R(i-j+1) - R(i-j-1)\right] = -2\left[R(j) - R(j-1)\right], \quad 1 \le j \le p \quad (13)$$

If the Z-transform is applied in both terms of (10) and (11), the following equations are obtained:

$$H^+(z) = \frac{E^+(z)}{X(z)} = \sum_{i=0}^{p} \frac{z^{-i}}{2}(h_i^+ + h_{i+1}^+) \quad (14)$$

$$H^-(z) = \frac{E^-(z)}{X(z)} = \sum_{i=0}^{p} \frac{z^{-i}}{2}(h_i^- - h_{i+1}^-) \quad (15)$$

In above equations, $h_0^+ = h_0^- = 2$ and $h_{p+1}^+ = h_{p+1}^- = 0$ are further introduced in comparison with those obtained by (12) and (13). In [2] it is shown that the LSF parameters are obtained by using the phase angle of the roots of polynomials $H^+(z)$ and $H^-(z)$. For instance, if p is even, LSF_1, LSF_3,…,LSF_{p-1} are computed by using $H^+(z)$ and LSF_2, LSF_4,…,LSF_p are computed by using $H^-(z)$. That is, computing the coefficients h_i^+ and h_i^- represents an alternative way to the traditional LSF computing that uses the symmetric and antisymmetric polynomials $P(z)$ and $Q(z)$, that are built based on the coefficients a_i [2], [3].
On the other side, the two time transforms that define $x^+(n)$ and $x^-(n)$ have the following Z-transforms:

$$\frac{X^+(z)}{X(z)} = \frac{z+1}{2} \quad (16)$$

and

$$\frac{X^-(z)}{X(z)} = \frac{1-z}{2}. \quad (17)$$

By examining the equations (16) and (17) it follows that time-transform that forms the average signal represents a low-pass filter and time transform that forms the differentiated signal represent a high-pass filter. In both cases, the cutoff frequency is $f_s/4$, because $|X^+(z)/X(z)|=|X^-(z)/X(z)|=0.707$ for $z=\exp(j\omega)$, $\omega=\pi f/(f_s/2)$ and $f=f_s/4$, f_s being the sampling frequency. That means that $x(n)$ is estimated by its previous samples that was first low-pass filtered, or respectively, high-pass filtered. Based on these statements, the linear prediction expressed by (8) will be called low-pass linear prediction and that expressed by (9), high-pass linear prediction. It follows that the minimum error energy for both linear predictions depends on spectral components of the sequence $x(n)$. Namely, for signals which have their energy concentrated on lower frequencies, the error for low-pass linear prediction is lower because the low frequency components are emphasized in comparison with higher frequency components. Similarly, for signals which have their energy concentrated on higher frequencies, the error for high-pass linear prediction is lower because the high frequency components are emphasized in comparison with lower frequency components.

III. COMPUTING THE MINIMUM ERROR ENERGY FOR LOW-PASS AND HIGH-PASS LINEAR PREDICTIONS

In the following we want to achieve a comparison between the three types of linear predictions (conventional, low-pass and high-pass) depending of the value of minimum error energy.
Therefore, the expressions of the minimum error energy for low-pass and high-pass linear predictions have to be computed. First, we compute this quantity for low-pass prediction. For this purpose, squared value of expression (10), and then, the equations (12) that allow obtaining the coefficients h_i^+, are used. Thus, the value of the error energy is

$$E_p^+ = \sum_n \left(x(n) + \sum_{i=1}^{p} h_i^+ x^+(n-i) \right)^2 = \sum_n \left(x(n) + \sum_{i=1}^{p} h_i^+ \frac{x(n-i) + x(n-i+1)}{2} \right)^2$$

$$= \sum_n \left(x^2(n) + \sum_{i=1}^{p} \frac{h_i^{+2}}{4}\left(x(n-i) + x(n-i+1)\right)^2 + x(n)\sum_{i=1}^{p} h_i^+ \left(x(n-i) + x(n-i+1)\right) \right.$$

$$+\sum_{i=1}^{p}\sum_{j=i+1}^{p}\frac{h_i^+h_j^+}{2}\left(x(n-i)+x(n-i+1)\right)\times\left(x(n-j)+x(n-j+1)\right)\Bigg)$$

$$\begin{aligned}&= R(0)+\sum_{i=1}^{p}\frac{h_i^{+2}}{4}\left(R(0)+R(0)+2R(-1)\right)\\&+\sum_{i=1}^{p}h_i^+R(i)+\sum_{i=1}^{p}h_i^+R(i-1)\\&+\sum_{i=1}^{p}\sum_{j=i+1}^{p}\frac{h_i^+h_j^+}{2}\left(R(-i+j)+R(-i+j+1)\right.\\&\qquad\left.+R(-i+j-1)+R(-i+j)\right).\end{aligned}\tag{18}$$

The second term and the last one from previous equation are grouped together and the resulting term is denoted by T. In the Appendix is demonstrated that if the coefficients h_i^+ and h_j^- are chosen such to minimize the expression of error energy, a simplified expression of T is:

$$T=-\sum_{i=1}^{p}\frac{h_i^+}{2}\left(R(i)+R(i-1)\right). \tag{19}$$

Replacing this expression of T in (18) the following expression of $E_{p+,\min}$ is obtained.

$$E_{p+,\min}=R(0)+\sum_{i=1}^{p}\frac{h_i^+}{2}\left(R(i)+R(i-1)\right). \tag{20}$$

In a similar mode, the expression for the minimum value of the error energy for the low-pass prediction can be obtained as

$$E_{p-,\min}=R(0)+\sum_{i=1}^{p}\frac{h_i^-}{2}\left(R(i)-R(i-1)\right) \tag{21}$$

It must be remarked that the expressions of minimum error energy (5), (20) and (21) are computed for values of n in range $0,\ldots,N-1,\ldots,N+p-1$, because the estimated signal can be computed in this range.

IV. EXPERIMENTAL RESULTS

In order to achieve the experiments, frames of speech signal from TIMIT data base [4], in English, as well as several frames acquired by a sound card of a computer, in Romanian, have been used. Also, frames from a synthesized signal with 5 components have been used. The last choice has the advantage that the five frequencies could be suitable chosen in order to cover the range of the spectrum. In all these cases, the sampling frequency was 8 kHz. The length of each frame was 200 samples or 25 ms. A conventional linear prediction of order p=10 was implemented first, by using (4), in order to compute the coefficients a_i. Then, the coefficients h_i^+ and h_i^- was computed for p=10, by using (12) and (13), respectively. Thus, LSF parameters were computed by using both the roots of polynomials $P(z)$ and $Q(z)$, and the roots of polynomials $H^+(z)$ and $H^-(z)$, respectively. The same values are obtained by each of two methods.

For each frame of signal, the LPC spectrum [1] and the value of minimum error energy for the three types of predictors were computed. In practice, the length of the error sequence $e(n)$ has to be equal to the length of the input sequence $x(n)$, and therefore the values of the minimum error energy were computed for n having values in range $0,\ldots N-1$.

First, two frames of synthesized signal were analyzed. The two signals contain the following frequencies, each of them having the same amplitude: 850 Hz, 1200 Hz, 2000 Hz, 3300 Hz, 3800 Hz, and 350 Hz, 900 Hz, 2000 Hz, 2700 Hz, 3100 Hz, respectively. Fig. 1 a) presents the LPC

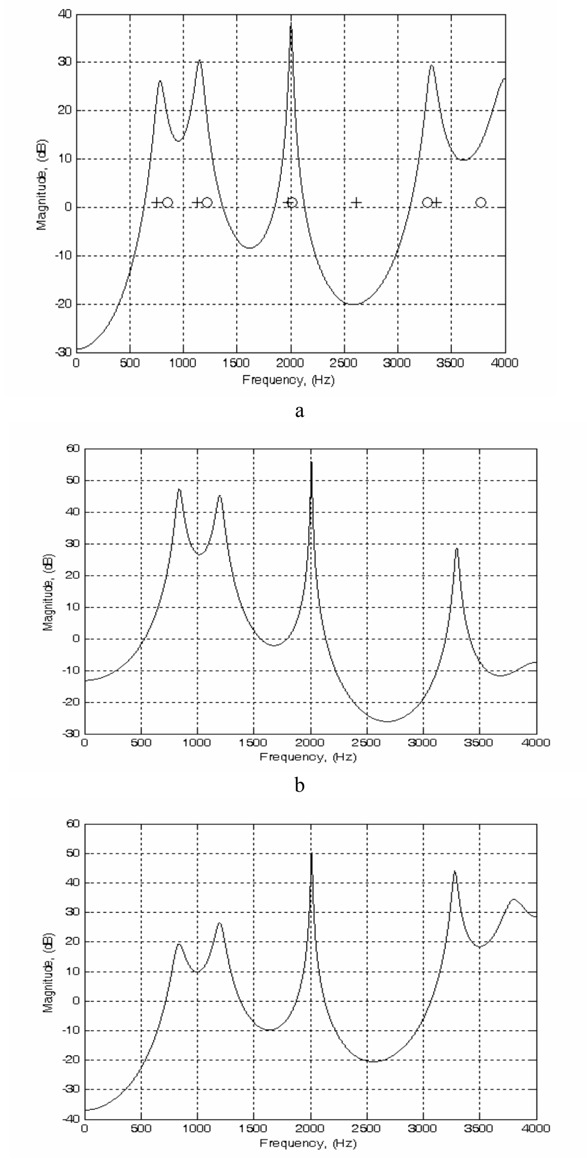
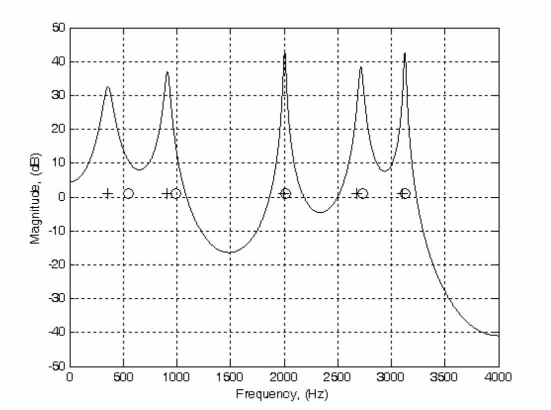
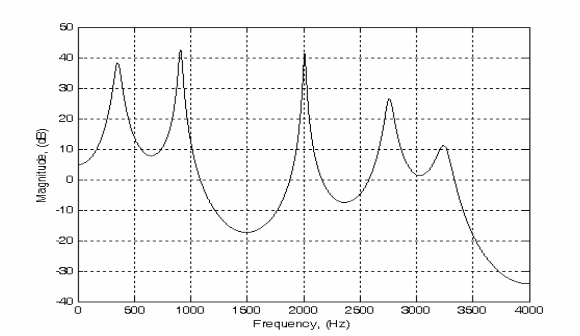


Fig.1. LPC spectrum for the first synthesized signal and its low-pass and high-pass versions

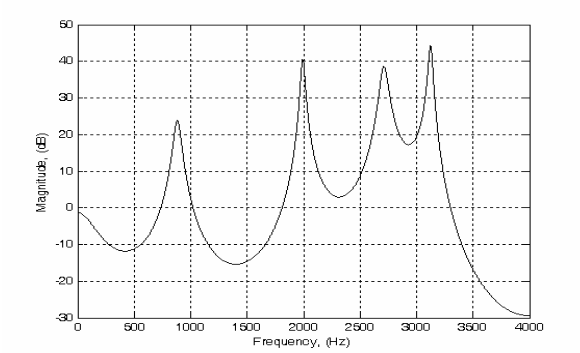
spectrum of the first signal, and the LSF parameters: those that have been computed by using roots of H^+ are represented with '+' symbol and those computed by using roots of H^- are represented by 'O' symbol. A cluster of 2 or 3 LSF's characterizes a peak (a formant) in the LPC spectrum. The width of the format depends on the closeness of the corresponding LSF's. A singular LSF characterizes a valley in the LPC spectrum. In the following figures, these LSF's property can be seen too. Fig. 1b) and 1c) represent the spectrum of low-pass filtered signal and high-pass filtered signal, respectively. Fig. 2 presents the same quantities for the second signal. Table 1 presents the value of the three errors for both the signals.



a



b



c

Fig.2. LPC spectrum for the second synthesized signal and its low-pass and high-pass versions

The first signal has an important part of the spectrum at higher frequency. After the high-pass filtering this part is emphasized, and instead after the low-pass filtering it is reduced. Thus, the energy error for high-pass prediction is much lower than that of low-pass prediction. Instead, the second signal has the main part of the spectrum at lower frequency and thus, the energy error for low-pass filtering is the lowest. In both cases the error energy for conventional prediction is much higher than the lowest of two other errors.

Table 1 Values of the minimum energy error for synthesized signals

	Signal 1	Signal 2
$E_{p+,\min}$	298	1.37
$E_{p-,\min}$	5.02	129
$E_{p,\min}$	315	201

In figures 3 to 6 are presented the LPC spectrum together with LSF parameters for four different frames of speech. The first two are frames of two different female speakers and last two are frames of two different male speaker. Table 2 presents the three values of the minimum error energy for the frames presented in fig. 3 to 6.

Table 2. Values of the minimum energy error for frames of speech signal

	Signal from fig.3	Signal from fig. 4	Signal from fig. 5	Signal from fig. 6
$E_{p+,\min}$	0.0456	0.0045	$4.5\cdot10^{-5}$	$3.6\cdot10^{-5}$
$E_{p-,\min}$	0.0183	0.0063	0,0098	0.452
$E_{p,\min}$	0.101	0.0532	0,0082	0.0253

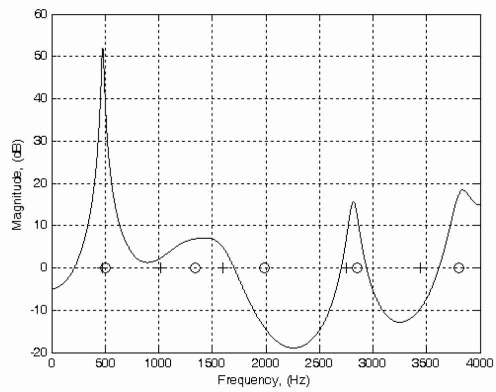


Fig.3 LPC spectrum for a female vowel [u:]

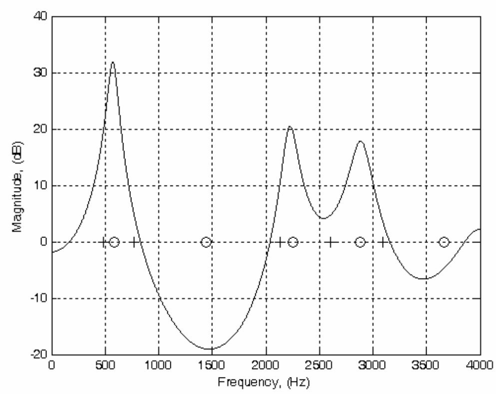


Fig.4 LPC spectrum for a female vowel [e:]

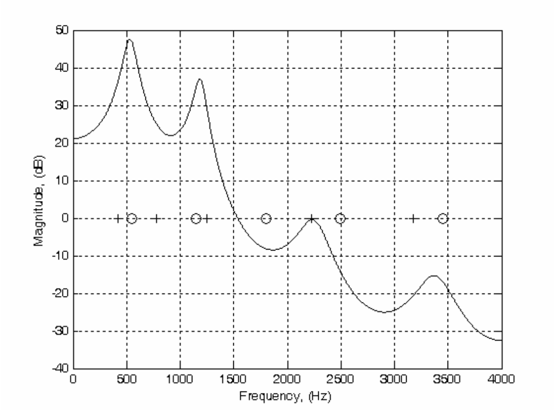


Fig.5 LPC spectrum for a male vowel [o:]

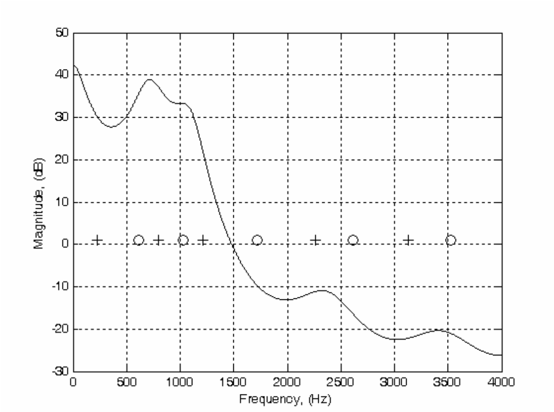


Fig.6 LPC spectrum for a male vowel [a:]

Based on these results, the following conclusions can be presented.

In the frames of female speakers, the LPC spectrum has components at both lower and higher frequencies. It follows that the errors of low-pass and high-pass predictions are close to each other, the error of conventional prediction being the biggest.

In the frames of male speakers, the LPC spectrum has the most important components at lower frequencies. It follows that the error of low-pass prediction is the lowest and then follows the error of conventional prediction, the error of high-pass prediction being the biggest.

V. CONCLUSIONS

The paper presents an important property of two new types of linear predictions previous introduced in [2]. Namely, for each of two predictors, the minimum energy error depends on spectral components of the analyzed frame. At least one of the two errors is much lower than the error in conventional prediction. These statements are valid for any type of signal and are based on the experimental results. Also in [5] is experimentally demonstrated that low pass linear prediction outperforms conventional linear prediction to finding the formants of the vowels speech data. Therefore, in the future, a demonstration of these statements should be achieved. In practice, this important property could be applied as follows. Both types of linear predictions can be applied to each frame of speech signal. Then, the prediction that allows obtaining the lowest value of the minimum error energy is chosen, and the prediction error sequence (together with the LSF parameters) is sent to receiver. Thus, fewer bits are necessary in order to quantize the prediction error sequence and a further compression could be obtained.

APPENDIX

In this section it is derived the expression of term T, that was used in section 3. Thus,

$$
\begin{aligned}
T &= T_1 + T_2 \\
&= \sum_{i=1}^{p} \frac{h_i^{+2}}{4}\big(R(0)+R(0)+2R(-1)\big) \\
&\quad + \sum_{i=1}^{p}\sum_{j=i+1}^{p} \frac{h_i^+ h_j^+}{2}\big(R(-i+j)+R(-i+j+1) \\
&\qquad + R(-i+j-1)+R(-i+j)\big).
\end{aligned}
$$

Each term of the sum in T_2 can be written twice if its denominator is considered 4. In this way, the computation of T is presented at the top of the next page. It can be seen that each term within squared brackets in (23) represents one of the equations used for computing the h_i^+ coefficients, (12). Then, (23) can be written as follows

$$
\begin{aligned}
T &= \frac{h_1^+}{4}(-2)\big(R(0)+R(1)\big) + \frac{h_2^+}{4}(-2)\big(R(1)+R(2)\big) \\
&\quad + \ldots + \frac{h_p^+}{4}(-2)\big(R(p-1)+R(p)\big) \\
&= -\frac{1}{2}\sum_{i=1}^{p} h_i^+\big(R(i-1)+R(i)\big).
\end{aligned}
$$

REFERENCES

[1] J. Makhoul "Spectral Linear Prediction-Properties and Applications", *IEEE Transactions on Acoustics Speech and Signal Processing*, vol.23, no. 3, 1975, pp. 283-296.

[2] T. Backstrom, P. Alku, T. Paatero, and B. Kleijn "A Time – Domain Interpretation for the LSP Decomposition", *IEEE Transactions on Speech and Audio Processing*, vol.12, no. 6, 2004, pp. 554-559.

[3] T. Backstrom, C.Magi and P. Alku, "Minimum Separation of Line Spectral Frequencies", *IEEE Signal Processing Letters*, vol.14, no. 2, 2007, pp. 145-147.

[4]"The DARPA TIMIT Acoustic Continuous Speech Corpus," National Institute of Standards and Technology (NIST), Gaithersburg, MD, 1990

[5] P.Alku and T. Backstrom, "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, 2004, pp. 93-99.

$$T=\frac{h_1^{+2}}{2}\left(R(0)+R(-1)\right)+\frac{h_1^+h_2^+}{4}\left(2R(1)+R(2)+R(0)\right)+\ldots+\frac{h_1^+h_p^+}{4}\left(2R(p-1)+R(p)+R(p-2)\right)$$

$$+\frac{h_1^+h_2^+}{4}\left(2R(1)+R(2)+R(0)\right)+\frac{h_2^{+2}}{2}\left(R(0)+R(-1)\right)+\ldots+\frac{h_2^+h_p^+}{4}\left(2R(p-2)+R(p-1)+R(p-3)\right)$$

....

$$+\frac{h_1^+h_p^+}{4}\left(2R(p-1)+R(p)+R(p-2)\right)+\frac{h_2^+h_p^+}{4}\left(2R(p-2)+R(p-1)+R(p-3)\right)+\ldots+\frac{h_p^{+2}}{2}\left(R(0)+R(-1)\right) \tag{22}$$

$$=\frac{h_1^+}{4}\left[2h_1^+\left(R(0)+R(-1)\right)+h_2^+\left(2R(1)+R(2)+R(0)\right)+\ldots+h_p^+\left(2R(p-1)+R(p)+R(p-2)\right)\right]$$

$$+\frac{h_2^+}{4}\left[h_1^+\left(2R(1)+R(2)+R(0)\right)+2h_2^+\left(R(0)+R(-1)\right)+\ldots+h_p^+\left(2R(p-2)+R(p-1)+R(p-3)\right)\right]$$

...

$$+\frac{h_p^+}{4}\left[h_1^+\left(2R(p-1)+R(p)+R(p-2)\right)+h_2^+\left(2R(p-2)+R(p-1)+R(p-3)\right)+\ldots+2h_p^+\left(R(0)+R(-1)\right)\right] \tag{23}$$