iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: V Month of publication: May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34680

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

Warped Linear Predictive Coding of Speech Signal of Processing

P. Choubey
Assistant Professor, Department of Physics, Govt. P.G. College Tikamgarh (M.P.)

Abstract: The analysis of speech signal in warped linear predictive coding is presented here. The auto-correlation function method is employed for analysis-power spectral density prediction gain and prediction coefficient have been studied for different adaptive parameters. The proposed system is found suitable in speaker identification. It is also suitable in speaker verification and speech recognition.

I. INTRODUCTION

Linear predictive coding was introduced 54 years ago [1] for speech signal processing and is the main tool uptil now. The main advantage of this coding is that it is confined to the all-pole characteristic of vowel spectra and spectral poles [2], Linear predictive coding is a well-established technique for speech compression at low bit rates and it has been applied in audio signal [3-6]. The warped linear predictive coding modifies the spectral representation. In this technique, unit delay elements in LPC are replaced by first order all pass filter. It is useful in the application of wide band speech and audio coding since it is connected clearly to the frequency resolution of human hearing. A Group of researchers [7-12] have employed WLPC techniques in the speech analysis and coding applications. All conventional techniques for parametric spectral estimation and linear filtering can be warped in a straight forward way. WLPC and LPC were compared in listening tests at various sampling rate and as a function of model order. This paper organized as Section 2, contains discussion of warped linear prediction filters. Correlations for WLPC are discussed in section 3. Section 4 contains lattice analysis for WLPC. Numerical results and discussion are given in section 5.

II. WLPC FILTERS

Following the notation [13, 14] the transfer function of first order all pass filter is given by

$$D(z) = \frac{z^{-1}-\lambda}{1-\lambda z^{-1}}$$

Where $\lambda \geq 0$ is a parameter and is related to sampling frequency f_s for a certain value of λ, the frequency mapping shows the similarity with the frequency of human auditory system. The parameter λ can be adjusted according to test signals for best results. So, it is called an adaptive parameter in a coder.

The output by first order all pass filter D(z) is given by

$$\hat{X}(z) = \left[\sum_{k=1}^{p} a_k D(z)^k\right] X(z) \qquad \text{.... (2)}$$

So, $$\hat{x}_n = \sum_{k=1}^{p} a_k d_k \left[x(n)\right] \qquad \text{.... (3)}$$

Where d_k [x(n)] = h(n) * h(n) ……… h(n) * x(n)

Where * sign indicates convolution and h(n) is the impulse of D(z). k is the order of prediction filter.

The predicted value of signal x_n is given as

$$\hat{x}_n = \sum_{k=1}^{p} a_k x_n D(z)^k \qquad \text{....(4)}$$

Where a_1, a_2, a_3,….a_p are prediction co-efficients.

Substituting the value of D(z) from eqn. (1) into eqn. (4), we get

$$\hat{x}_n = \sum_{k=1}^{p} a_k x_n \left\{\frac{z^{-1}-\lambda}{1-\lambda z^{-1}}\right\}^k \qquad \text{....(5)}$$

For a 1st order prediction filter, putting k = 1 and solving eqn. (5), we get

$$\hat{x}_n = a_1 x_{n-1} - \lambda a_1 x_n - \lambda^2 a_1 x_{n-1} \quad \text{….(6)}$$

For $\lambda = 0$, $\hat{x}_n = a_1 x_{n-1}$ is the 1st order prediction for LPC.

For a 2nd order filter, putting k = 2 solving eqn. (5), we get

$$\hat{x}_n = a_1 x_{n-2}\left(1-\lambda^2\right) - \lambda a_1 x_n + \lambda^2 a_2 x_n - 2\lambda a_2 x_{n-1} + a_2 x_{n-2}\left(1 - 4\lambda^2 + \lambda^4\right) \quad \text{….(7)}$$

In similar fashion, we can determine higher order of filters.

The output sequence of the all pole model satisfies the difference equation

$$\hat{x}_n = \sum_{k=1}^{p} a_k x_n D(z)^k + GV_{in} \quad \text{….(8)}$$

Where G is prediction gain and V_{in} is the input sequence.

$$E[GV_{in}]^2 = G^2 E[V_{in}^2]$$

$$E\{V_{in}^2\} = 1$$

Prediction error filter is given by

$$A(z) = 1 - \sum_{k=1}^{p} a_k D(z)^k \quad \text{….(9)}$$

So, gain term of the filter

$$G = A(-\lambda) = 1 - \sum_{k=1}^{p} a_k (-\lambda)^k \quad \text{…. (10)}$$

III. AUTO-CORRELATION AND POWER SPECTRAL DENSITY

The auto-correlation function is defined as

$$r(k) = E\left[x_i, x_i \left\{ \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \right\}^k \right] \quad \text{....(11)}$$

Where x is the signal.

From eqn. (11), we get the relation between correlation function r(k) for WLPC and the correlation function $\phi(k)$ for LPC.

For example,

r(1) =

$= \phi(1) - \lambda\phi(0) - \lambda^2\phi(1)$ ….(12)

$r(2) = \phi(2)(1 - 4\lambda^2 + 3\lambda^4) - 2\phi(1)(\lambda - \lambda^3) - \lambda^2\phi(0)$ ….(13)

$r(3) = \phi(3)(1 - 9\lambda^2 + 15\lambda^4) - \phi(2)(3\lambda - 9\lambda^3 + 6\lambda^5) + \phi(1)(3\lambda^2 - 3\lambda^4) - \lambda^3\phi(0)$ ….(14)

and other (not shown)

Where, $\phi(k) = \frac{1}{N}\sum_{i=1}^{N} x_i x_{i+k}$; k = 0, 1, 2, ……….p ....(15)

The normalized prediction error is given by the relation

$$e = 1 - \sum_{k=1}^{p} a_k \frac{r(k)}{r(0)} \quad \text{....(16)}$$

The power spectral density is determined by Z-transform of auto-correlation function.

$$S(z') = \sum_{n=1}^{N} r(n) z^{-n} \quad \text{….(17)}$$

In terms of frequency

$$S(\omega) = \sum_{n=1}^{N} r(n) e^{-j\omega n}$$

The normalized cress sectional area of vocal tract is given as

$$A_{N'-1} = A_{N'} \frac{1+K'_{N'-1}}{1-K'_{N'-1}} \qquad \text{....(18)}$$

Where N' = 1, 2, ……7 are number of cross section.

$A_{N'}$ is cross section

$K'_{N'}$ is reflection co-efficient for WLPC

IV. LATTICE ANALYSIS FOR WLPC

The two recursive equations for lattice analysis are given as

$$f_{m+1}(n) = f(n) - k'_{m+1} b_m D(z) \qquad \text{....(19)}$$

and

$$b_{m+1}(n) = b_m D(z) - k'_{m+1} f_m(n) \qquad \text{...(20)}$$

Where $f_m(n)$ and $b_m(n)$ are forward and backward prediction error of the m^{th} stage lattice filter at time n respectively.

Autocorrelation and cross-correlation functions for the m^{th} stage can be defined as

$$\alpha_m^f(t) = \frac{1}{N}\sum_{n=0}^{N-1} f_m(n)\, f_m(n-t) \qquad \text{....(21)}$$

and

$$\alpha_m^b(t) = \frac{1}{N}\sum_{n=0}^{N-1} b_m(n)\, b_m(n-t) \qquad \text{....(22)}$$

$$\beta_m^b(t) = \frac{1}{N}\sum_{n=0}^{N-1} f_m(n)\, b_m(n-t) \qquad \text{....(23)}$$

and

$$\beta_m^b(t) = \frac{1}{N}\sum_{n=0}^{N-1} b_m(n)\, f_m(n-t) \qquad \text{....(24)}$$

Substituting eqn. (19) and (20) into (21), (22), (23) and (24), we have two recursive equations as

$$a_{m+1}(t) = a_m(t)\left[1 + \left(k'_{m+1}D(z)\right)^2\right] k'_{m+1}D(z)\left[\beta_m(t+1) + \beta_m(t-1)\right) \qquad \text{...(25)}$$

and

$$\beta_{m+1}(t) = \beta_m(t+1)\, D(z) - k'_{m+1}\, \alpha_m(t)\left[1 + (D(z)^2)\right] + k'^2_{m+1}\, D(z)\beta_m(t-1) \qquad \text{...(26)}$$

Where

and $$\left.\begin{array}{l} \alpha_m^f(-t) = \alpha_m^b(t) = \alpha_m(t) \\ \beta_m^f(-t) = \beta_m^b(t) = \beta_m(t) \end{array}\right\} \text{ for all t}$$

Optimal parcor – coefficients are given by

$$k'_{m+1} = \frac{\beta_m(1)}{\alpha_m(0)} \qquad \text{....(27)}$$

Hardware realization of lattice filter is given in fig. (1). The block diagram of speech signal transmitter and receiver are shown in fig. (2) and fig. (3) respectively.

V. NUMERICAL RESULTS AND DISCUSSION

Numerical calculation of correlation functions r(1), r(2), r(3), r(4), r(5), r(6), r(7) and r(8) have been performed for $\lambda = 0$, 0.1 and 0.2 for 100 samples from eqn. (11). Results are shown in fig. (4). From the figure, we infer that correlation function decreases sharply for prediction order k = 2, almost constant for $3 \geq k \geq 2$ and for $k > 3$, it again decreases slowly. Correlation function decreases with increase of λ for any prediction order.

Prediction co – efficient $a_0, a_1, a_2, a_3, a_4, a_5, a_6$ and reflection co – efficient $k'_1, k'_2, k'_3, k'_4, k'_5$ and k'_6 are compared using Levinson algorithm with the help of MATLAB for $\lambda = 0$, $\lambda = 0.1$, $\lambda = 0.2$, $\lambda = 0.3$ and $\lambda = 0.4$. The results so obtained, are depicted in fig. (5).

Normalized prediction error is computed using eqn. (16) for prediction order k = 1, 2, 3, 4, 5 and 6 for $\lambda = 0$, 0.1 and 0.2. The results are shown in fig. (6). From the figure, we conclude that prediction error (e) decreases with the increase of prediction, order (k). As λ increases, the prediction error also increases for any prediction order. The prediction error for $\lambda = 0.1$ and 0.2, decreases sharply from order k = 1 to 2, and then it decreases very slowly between k = 2 to 6, whereas the curve for $\lambda = 0$ indicate much faster decrease as compared to others.

Normalized cross sectional area of vocal tract is numerically computed using eqn. (18) with the help of reflection co-efficient. A_7 is calculated by taking the average area of lips for the pronunciation of letter 'a' by different speakers. A graph is plotted to show the variation of normalized cross sectional area with distance from glottis in different sections for $\lambda = 0$, 0.1 and 0.2. The results have been compared with the results of author [15]. Our results are indicated in fig. (7). The curve for $\lambda = 0.1$ gives the results close to author [15]. Here, total length of vocal tract l = 15 cm, the sampling period $T = 2\tau$ where τ is the one way propagation time in a single section. If there are N sections $\tau = \frac{1}{cN}$ where c is the velocity of wave in the tube.

Here, c = 21,000 cm/sec

$$\frac{1}{2T} = \frac{1}{4\tau} = \frac{Nc}{4l} = \frac{N}{2}(700)H_Z$$

This implies that there will be about $\frac{N}{2}$ resonances per 700 Hz of frequency for a vocal tract of total length 15 cm. If $\frac{1}{T} = 4900$ Hz, N = 7, Thus, number of sections are 7.

Power spectral density, $S(\omega)$ for digital angular frequency CD for λ =0.5, 0.6 and 0.7 is computed using MATLAB and shown in fig.(8).

Prediction gain of the filter has been calculated using eqn. (10) for $\lambda = 0.6$, $\lambda = 0.7$ and $\lambda = 0.8$ for prediction order 1, 2, 3, 4, 5 and 6. The results are produced in fig. (9). The prediction gain first increases for k from 2 to 3, then decreases for k from 3 to 4 and finally increases for higher values of k. With the increase of adaptive parameter A, prediction gain increases at any prediction order. The dotted line indicate the experimental values of predictive gain due to P. Noll [16].

VI. CONCLUSION

The auto correlation functions are found to be different for different speakers. Thus, auto-correlation function at a particular prediction order has different values for different speakers.

The warped linear prediction coding covers the range of values of auto correlation functions with the help of adoptive parameters. Values of auto-correlation can be adjusted for a particular speaker at fixed prediction order. The prediction gain of filter also indicates variation at a particular prediction order due to different speakers. The variation of prediction gain at certain prediction order can be achieved by adjusting the adaptive parameter. In this way the transmission of speech signals by different speakers is possible with the help of WLPC. The warped frequency which can be obtained by the adjustable adaptive parameter can be employed in the transmission of signal for a particular speaker. The technique of frequency warping can be employed in speaker verification and speech recognition. Warped linear predictive coding is suitable for both the high sampling rate low sampling rate. At low sampling rate typically at 80 KHz/sec, conventional LPC and WLPC, produce approximately the similar results but the difference between LPC and WLPC results are significant at high sampling rate. All types of speech sound i.e. voiced and unvoiced speech requires sampling rate greater than 20 KHz/sec, to exactly represent the signals. Thus, WLPC is suitable for transmission of both the voiced and unvoiced speech signals.

A. Captions To Figures

1) Hardware realization of lattice filter;
2) Block diagram of WLPC transmitter.
3) Block diagram of WLPC receiver.
4) Variation of correlations with prediction order for λ = 0, 0.1, 0.2.
5) Values of prediction co-efficients and reflection co-efficients for λ = 0, 0.1, 0.2, 0.3, and 0.4.
6) Variation of prediction error with prediction order for λ = 0, 0.1, 0.2.
7) Normalized cross-section area versus distance from glottis for λ = 0, 0.1, 0.2.
8) Change of power spectral density with digital angular frequency for λ= 0.5, 0.6, 0.7
9) Prediction gain versus prediction order for λ = 0.6, 0.7, 0.8.

REFERENCES

[1] B.S. Atal and M.R. Schroeder, "Predictive coding of speech signals", In Proc. IEEE Conf. on Commu. and Processing (1967) pp.360-361.
[2] M.R. Schroeder, "Linear Prediction, extremal entropy and Prior information in speech signal analysis and synthesis", Speech Commu., Vol, No. 1 (May 1982) pp. 9-20.
[3] A. Harma, "Linear predictive coding with modified filter structures", IEEE Trans. Speech Audio Processing, Vol. 9 (Nov. 2001) pp.769-777.
[4] Paavo Alku and Tom Backstrom, "Linear predictive method for improved spectral modeling of lower frequencies of speech with small prediction orders", IEEE Trans. on Speech and Audio Processing, Vol. 12, No.2 (March 2004) pp. 93-98.
[5] Mi. Suk Lee, Hong Kook Kim and Hwang SooLee, "LPC analysis incorporating spectral interpolation for speech coding", Electronics Letters, Vol. 35, No. 3 (4th February 1999) pp. 200-201.
[6] Bessette, R. Salami, C. Laflamme and R. Lefebvre, "A Wideband Speech and Audio Codec at 16/24/32 k bit/s using hybrid ACELP/TCX techniques", In Proc. IEEE Workshop Speech Coding, Porvoo, Finland (1999).
[7] U.K. Laine M. Karjalainen and T. Altosaar, "WLPC in speech and audio processing", In proc. IEEE Int. Conf. Acoust, Speech and Signal Proc., Vol. III (1994) pp. 349-352.
[8] S. Imai, "Capstral analysis synthesis on the mel frequency scale", In Proc., IEEE Int. Conf. Acoust., Speech, and Signal Proc., Boston (April 1983) pp.93-96.
[9] M. Deriche and D. Ning, "A Novel Audio coding scheme using Warped linear prediction Model and Discrete wavelet Transform", Audio, Speech, and Languge Processing, IEEE trans. on Vol. 14 (Nov. 2006) pp.2039-2048.
[10] M. Boillot and J. Harris, "A Warped Bandwidth Expansion Filter", Acoustics, Speech and Signal Processing, 2005. Proceeding. IEEE Int. Conf. on Vol. l (March 18-23, 2005) pp.65-68.
[11] Hauke Kruger and Peter Vary, "Warped linear prediction for improved perceptual Quality in the SCELP LOW DELAY AUDIO CODEC (W-SCELP)", Proc. of the 10th Int. Conf. on Digital Audio Effects (DAFx-07), Bordeacs, France (Sept. 10-15, 2007).
[12] T. Van Waterschoot and M. Moonen. "Adaptive feedback cancellation for Audio signals using a warped all-pole near-end signal model", Acoust., Speech and Signal Processing, IEEE Int. Conf. (March 31, April 2008) pp. 269-272.
[13] A. Harma, "Implementation of frequency-warped recursive filters", Signal Processing, Vol. 80, No.3 (February 2000) pp.543-548.
[14] Aki Harma and Unto K. Laine, "A comparison of warped and conventional linear predictive coding", TEEE Trans. On Speech and Audio Processing, Vol. 9, No. 5 (July 2001) pp.579-588.
[15] Thomas F. Quatieri, "Discrete-time speech signal processing", Published by Pearson Education (Singapore) Pvt. Ltd. (2002) pp. 205
[16] P. Noll, "A comparative study of various schemes for speech encoding", Bell System Tech. J., Vol. 54, No.9 (Nov.1975) pp. 1597-1614.

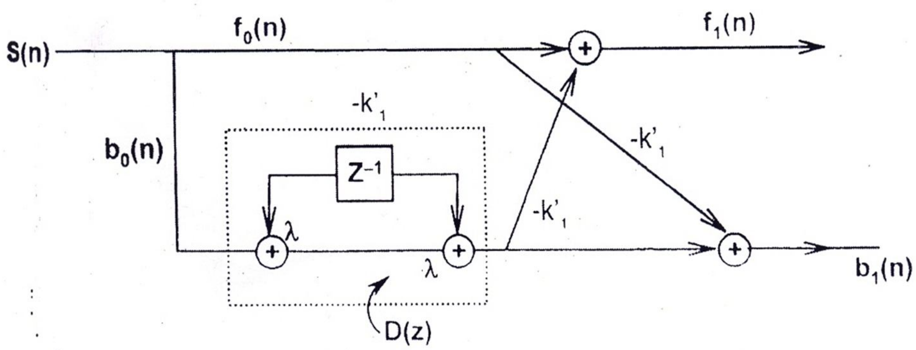

Fig. 1: Hardware realization of lattice filter

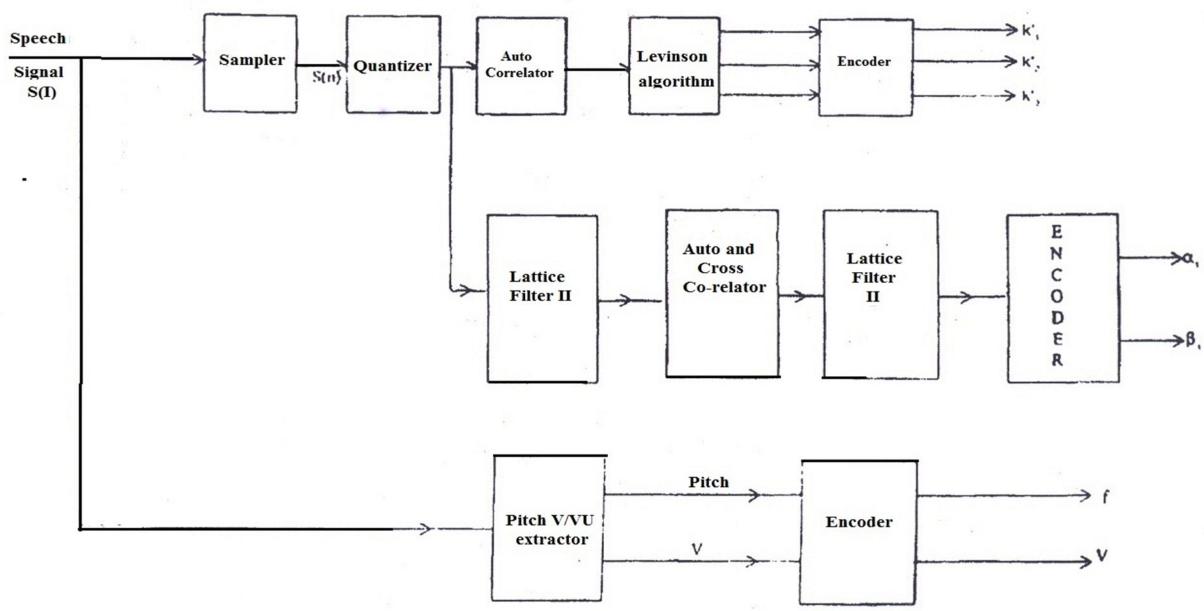

Fig. 2: Block diagram of WLPC transmitter

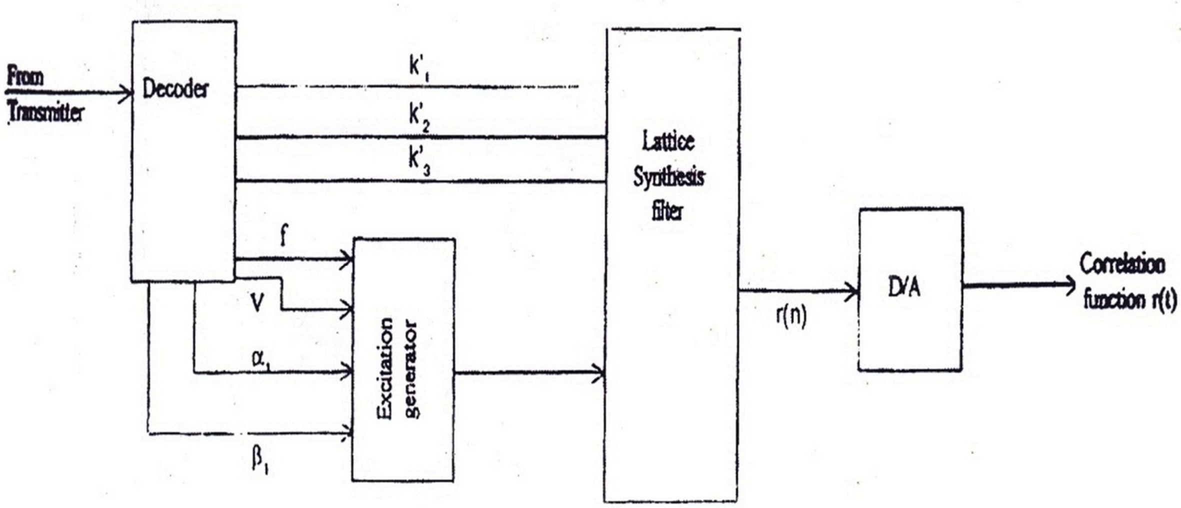

Fig. 3: Block diagram of WLPC receiver.

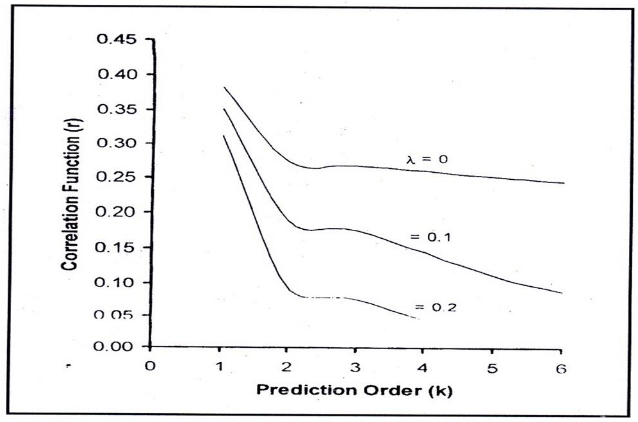

Fig. 4: Variation of correlations with prediction order for λ = 0, 0.1, 0.2.

For λ = 0
a = [1 - 0.2830 -0.2203 -0.1555 -0.1098 - 0.0806,-0.0540]
k' = -0.7115
-0.3920
-0.2391
-0.1511
-0.0961
-0.0540

For λ = 0.1
a = [1 -0.3562 -0.2725 -0.0566 -0.0272 -0.0044 0.0253]
k' = -0.5362
-0.2997
-0.0606
-0.0187
0.0046
0.0253

For λ = 0.2
a = [1 -0.2738 -0.1924 0.1313 -0.0334 -0.0230 0.0541]
k' = -0.3013
-0.1659
0.1162
-0.0253
-0.0082
0.0541

For λ =0.3
a = [1 0.0086 0.0019 0.3596 -0.1052 -0.1301 -0.0285]
k' = 0.0124
0.0486
0.3797
-0. 1063
-0.1299
-0.0285

For λ = 0.4
a = [1 -0.1679 0.0161 1.2921 -0.0825 0.3430 0.9779]
k' = 0.4458
0.6312
2.3766
-0.9830
11.6032
0.9779

Fig. 5: Values of prediction co-efficients and reflection co-efficients for λ = 0, 0.1, 0.2, 0.3, and 0.4.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

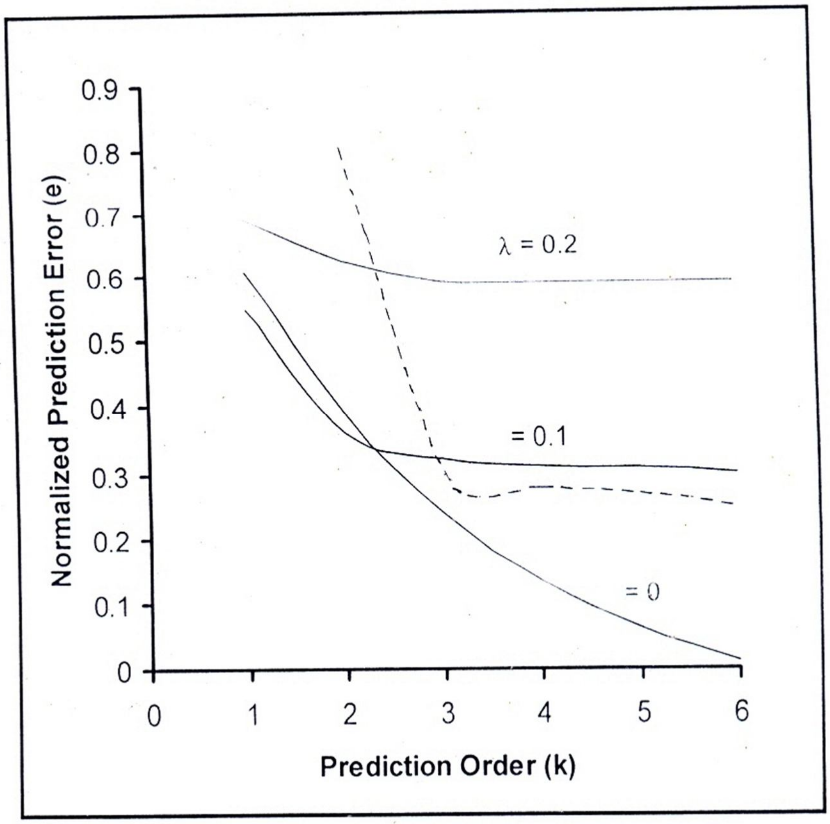

Fig. 6: Variation of prediction error with prediction order for λ = 0, 0.1, 0.2.

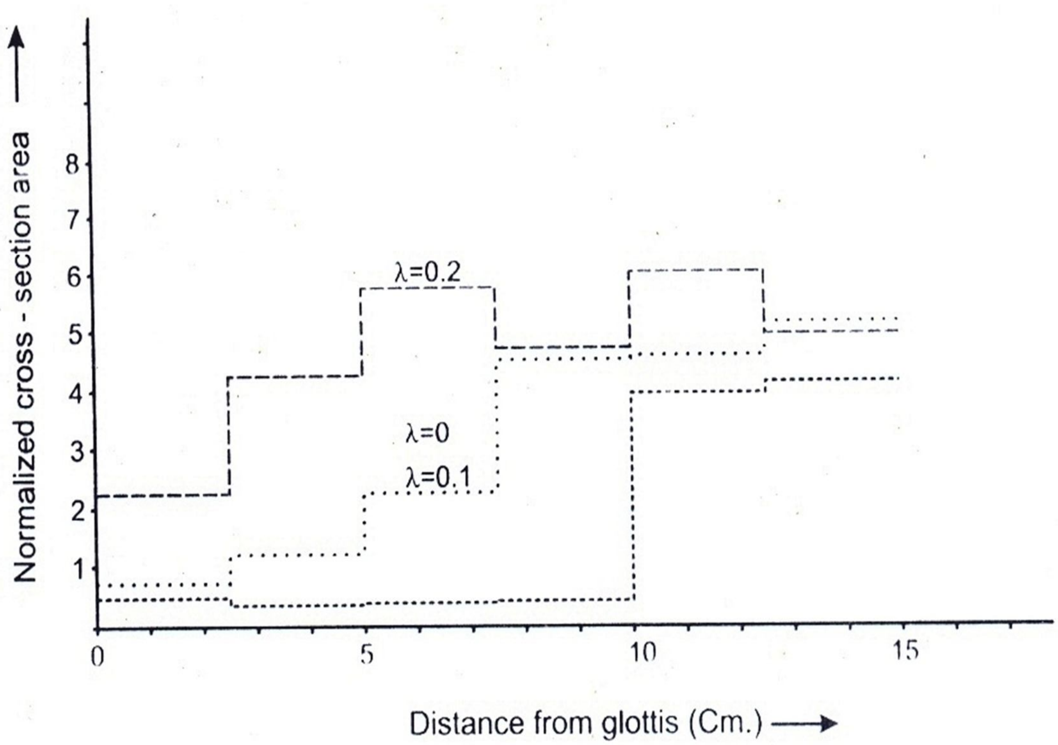

Fig. 7: Normalized cross-section area versus distance from glottis for λ = 0, 0.1, 0.2.

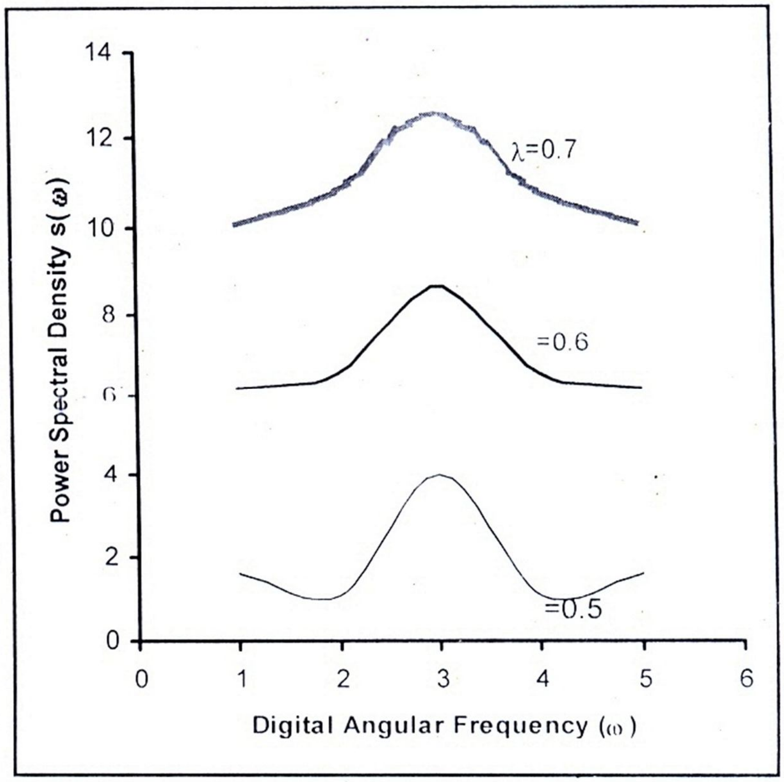

Fig. 8: Change of power spectral density with digital angular frequency for λ = 0.5, 0.6, 0.7

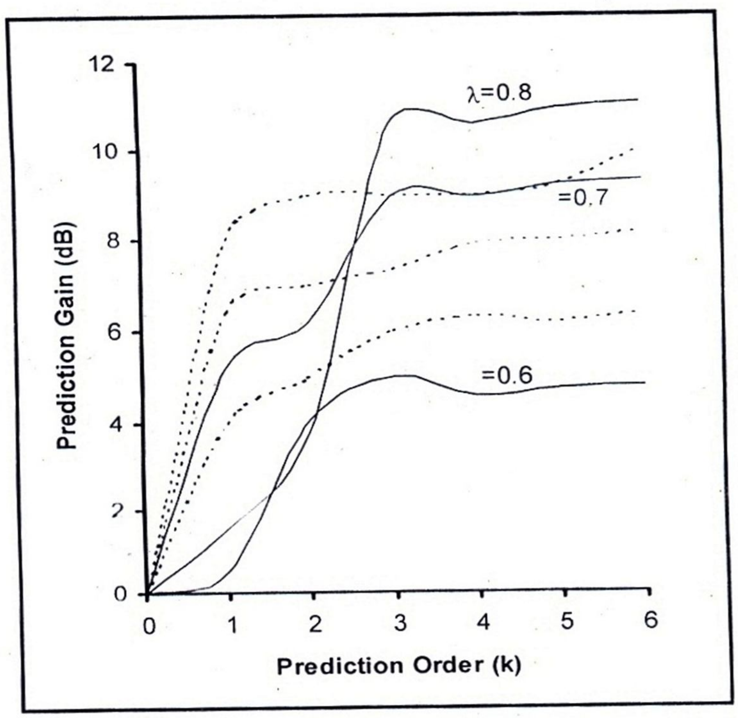

Fig. 9 : Prediction gain versus prediction order for λ = 0.6, 0.7, 0.8.