Nonlinear Prediction of Speech Signal using Volterra-Wiener Series

Hemant A. Patil and Tanvina B. Patel

Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar-382007, India

{hemant_patil, tanvina_bhupendrabhai_patel}@daiict.ac.in

10.21437/Interspeech.2013-31

Abstract

Linear Prediction (LP) analysis has proven to be very effective and successful in speech analysis and speech synthesis applications. This may be due to the fact that LP analysis captures *implicitly* the time-varying vocal tract area function. However, it captures only the *second-order* statistical relationships and only the *linear* dependencies in the sequence of samples of speech signals (and not the higher-order relations), as a result of which the LP residual is also *intelligible*. This paper studies the effectiveness of nonlinear prediction (NLP) of the speech signal by using the state-of-the-art Volterra-Wiener series and uses a novel chaotic titration method to analyze the chaotic characteristics of the residual obtained by both the LP and NLP methods. The experimental results demonstrate that the proposed NLP approach gives less prediction error, relatively flat residual spectrum, less PESQ score (i.e., objective evaluation of MOS to a certain extent) and less chaoticity than its LP counterpart. Finally, the L^1 norm and L^2 norm of NLP residual was found be *relatively less* than LP residual for five instances of voiced and unvoiced regions extracted from speakers of TIMIT database.

Index Terms: Linear prediction, nonlinear prediction, residual, Volterra-Wiener series, chaotic titration, noise limit

1. Introduction

The acoustic theory of speech production assumes the speech production mechanism to be a *linear* system. Linear Prediction (LP) analysis has proven to be very effective in speech analysis and speech synthesis applications. Extensive work has been carried out on LP analysis of speech [1], [2]. The effectiveness of LP analysis lies in its ability to capture *implicitly* the frequency response of the time-varying vocal tract area function.

In realistic scenarios, the speech production system is known to have some *nonlinearity* due to the nonlinear interaction or *coupling* of the source and the system. For example, the first formant (and higher formants to a certain extent [3]) of the vocal tract is known to interact *nonlinearly* with the glottal flow waveform; resulting in sudden drop of acoustic pressure at the lips during the opening phase of the glottis [4]. Not only the vocal fold collision, but also the nonlinear pressure in the glottis, delayed feedback of the mucosal wave, the nonlinear stress and strain curves of the tissues, the turbulence, and the pressure at the glottis; all contribute to the *nonlinearity* in the speech signal. In many applications, LP forms the basis of the *source-filter* model where the filter is constrained to be an all-pole linear filter. The assumption is that the frequency response of the vocal tract consists of only poles (i.e., all-pole model). However, such a model is not sufficient to analyze nasal and fricatives.

One of the first attempts to model nonlinearities (and nonlinear dependencies) in speech signal was reported by Thyssen *et. al.* [5] purely in the context of speech coding. This work proposed two methods for nonlinear prediction (NLP), *viz.*, second-order Volterra filter and on a time-delay neural network. One of the important limitation of this study was that predictor was applied twice (first for LP model and second time for NLP model) [6], [7]. There been interesting attempts to model the global behavior of Volterra-Wiener (VW) series using artificial neural networks [8], [9], [10]. In particular, the approach presented in [8] alleviates an important limitation of computational complexity associated with calculation of Volterra kernel by using back propagation training of higher-order nonlinear models in the three-layered perceptrons (TLP). However, none of these approaches explain how to compute the Volterra kernels. To address this issue, an interesting application was presented in [11] to model the nonlinear characteristics of electronic devices (such as diode capacitance curve). Thereafter, in [12], the instability of VW series non-linear predictive filter used for speech coding was studied and then a scheme was proposed that detects those frames for which, after stabilization, including the quadratic predictor is beneficial. Very recently, a novel application of Volterra series is presented to analyze the multilayered perceptron (MLP) to estimate the posterior probability of phoneme for automatic speech recognition (ASR) task [13]. In addition, in the same study, Volterra kernels are used to capture *spectro-temporal* patterns that are learned from the Mel filterbank log-energies (during training of system for each phoneme).

In this paper, we present an approach of predicting speech signal by nonlinear prediction model that exploits the *nonlinear* relationship among the sequence of samples of speech signal by the VW series approximation (containing linear and non-linear terms at the same time). In addition, we give evidence of nonlinearities in speech via Teager energy approach. The log-spectral distance (LSD) of LP and NLP residual is studied followed by the Perceptual Evaluation of Speech Quality (PESQ) score of the residuals to justify the lower Mean Opinion Score (MOS) of the NLP residual. Experimental results on L^1 norm and L^2 norm of voiced and unvoiced regions of speech data for *100* speakers of TIMIT database is discussed. Furthermore, a novel chaotic titration [14], [15] method has been implemented to study the chaotic behavior of speech and its LP and NLP residuals to visualize the methods in a common framework.

2. Linear *vs.* nonlinear prediction

This section presents the effectiveness of NLP model over LP model. The nonlinear prediction method is based on the Volterra-Wiener-Korenberg (VWK) series approximation.

2.1. Linear Prediction

Speech signal is produced by the *convolution* of excitation source and time-varying vocal tract system components. This excitation and vocal tract components are to be separated from the speech signal to study them independently for several speech analysis applications such as pitch detection and formant estimation. Next, the LP analysis is discussed in brief.

For a speech signal $s(n)$ of length N, the predicted signal $\hat{s}(n)$ can be represented using p predictor memory as follows,

$$\hat{s}(n) = -\sum_{k=1}^{p} a_k s(n-k), \quad (1)$$

where a_k's are the LP coefficients (LPC), and the prediction error $e_{LP}(n)$ or the LP residual can be computed (from (1)) as,

$$e_{LP}(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^{p} a_k s(n-k). \quad (2)$$

The LP methods for obtaining the *optimal* predictor coefficients are based on minimizing the l^2 energy of LP residual signal and it is given by,

$$E = \sum_{n=0}^{N-1} |e_{LP}(n)|^2 . \quad (3)$$

The LP residual has been used extensively for estimating the pitch period or epoch (i.e., glottal closure instants abbreviated as GCIs) of the speech segment [16] and is also used as *excitation source* for speech coding purpose [17].

2.2. Nonlinear prediction using Volterra polynomials

The set of Volterra functionals is *complete* (i.e., every Cauchy sequence converges to a *limit point* which belongs to the same *function* space) [18]. It means that every continuous functional of a signal $x(t)$ can be approximated with arbitrary precision as a sum of a finite number of Volterra functions in $x(t)$. This result was a generalization of the Weierstrass-Stone theorem (i.e., every continuous function of a variable $x(t)$ can be approximated with arbitrary precision as the sum of a finite number of polynomials in $x(t)$) [19], [20].

To perform the nonlinear prediction of speech, consider a dynamical system with input data series $x(n)$ and the output $y(n)$ at time instant $n=1,2,\ldots,N$, in multiples of sampling time τ. A power series expansion such as the Taylor series expansion may be used to describe the output of the system as,

$$y(t) = \sum_{p=0}^{\infty} c_p x^p(t). \quad (4)$$

A nonlinear system with k memory terms can be then represented by means of an extension of (4). This extension, known as the Volterra series expansion, which relates the input and output of the system is used. For a dynamical system, a closed-loop version of the Volterra series is used in which the output $y(n)$ feeds back as a delayed input (i.e., $x(n)\equiv y(n)$). Therefore, we analyze the univariate time series by using a discrete VW series of degree d and predictor memory k to calculate the predicted time series $\hat{y}(n)$ given by:

$$\hat{y}(n) = a_0 + a_1 y(n-1) + a_2 y(n-2) + \ldots + a_k y(n-k) + a_{k+1} y(n-1)^2$$
$$+ a_{k+2} y(n-1) \times y(n-2) + \ldots + a_{M-1} y(n-k)^d$$

$$\therefore \hat{y}(n) = \sum_{m=0}^{M-1} a_m z_m(n), \quad (5)$$

where the functional basis $\{z_m(n)\}$ is composed of all the distinct combinations of the *embedding space* coordinates up to degree d with a total dimension $M = (k+d)!/(k!d!)$ [14]. Thus, each model is parameterized by k and d corresponding to the *predictor memory* and the *degree* of nonlinearity in the model, respectively. The coefficients a_m's in (5) are estimated by Korenberg's fast algorithm using Gram-Schmidt procedure from linear and nonlinear autocorrelation of the data-series itself [21]. From (2) and (4), for NLP model, $s(n) = y(n)$ and $\hat{s}(n) = \hat{y}(n)$, Therefore, NLP residual, *viz.*, e_{NLP} is given by,

$$e_{NLP}(n) = y(n) - \hat{y}(n). \quad (6)$$

To analyze the prediction error by LP and NLP methods, we take a segment of speech from CMU-ARCTIC database [22], sampled at *32 kHz* and of size *400* samples. The signal is predicted for different polynomial degree d. The number of memory terms required for the prediction are kept constant, i.e., $k=3$. The energy in prediction error (3) was found to be *0.0216* (for $d=1$ and $k=3$, i.e., LP), *0.0193* (for $d=2$ and $k=3$, i.e., NLP), and *0.0169* (for $d=3$ and $k=3$, i.e., NLP). Thus, it is evident that the energy in prediction error (3) decreases as the degree of polynomial increases; it is due to the combinations of the larger number of previous samples (and their *nonlinear* dependencies) are taken into account to predict the speech.

In addition to this, to obtain an analysis on a large number of speakers, the TIMIT database [23] was chosen and the L^1 norm and L^2 norm of the residuals by LP model ($d=1$, $k=12$) and NLP model ($d=2$, $k=12$) for *five* instances of voiced (/aa/) and unvoiced (/s/) segments obtained each from *100* speakers of training part of TIMIT database were estimated. It can be observed from Table 1 that the L^1 and L^2 norm for NLP residual energy is less for both /aa/ and /s/ phonemes. In addition, the norm values for /aa/ segments are less than the /s/ segments, indicating that the unvoiced speech may be more difficult to predict than the voiced speech. The lower values for NLP norm indicate that NLP-based approach has predicted both kinds of speech signals more accurately than the LP-based approach. This may be due to the fact that more number of polynomials is used in NLP resulting in less error power.

Table 1. *The value of LP and NLP residual norm averaged over 100 speakers from TIMIT database* [23]

Phoneme	/aa/		/s/	
Method	(L^1norm)	(L^2 norm)	(L^1 norm)	(L^2 norm)
LP	3.753532	0.130202	6.001931	0.181783
NLP	3.179727	0.105471	5.812873	0.174548

Table 2. *The computation time in seconds for speech segment 'I have no idea' from CMU-ARCTIC database (arctic_a0057.wav)* [22], *Fs=32 kHz when predicted by LP and NLP method for different no. of polynomial terms.*

Method	Number of coefficients *(M=(d+k)!/k!d!)*							
	3	6	10	15	21	28	36	48
k for LP *(d=1)*	2	5	9	14	20	27	35	47
Time (sec)	1.2	2.2	3.8	6.2	12	24	45	95
k for NLP *(d=2)*	1	2	3	4	5	6	7	8
Time (sec)	1.17	2.18	3.6	5.45	9.38	17.4	29.8	42.3

As far as the computation time is considered, the prediction by NLP-based approach takes less amount of time than LP approach by VWK series for the same number of coefficients. This may be due to the fact that for same number of predictor coefficients, the NLP approach requires less memory terms k. This is shown in Table 2. The time calculated is rounded off to one decimal place and computed using *Core i5* processor with *4* GB RAM.

3. Analysis of LP residual *vs.* NLP residual

In this section, we study the behavior of Teager energy operator (TEO) profile and residuals (computed using LP and NLP methods) in the vicinity of GCIs in order to investigate the evidence for nonlinear dependencies in samples of the speech signal. Figure 1(c) shows the TEO profile of the speech segment having peaks around GCI pulses. In between the TEO pulses, *bumps* are observed which indicate the *nonlinearities* associated with the production of speech [4].

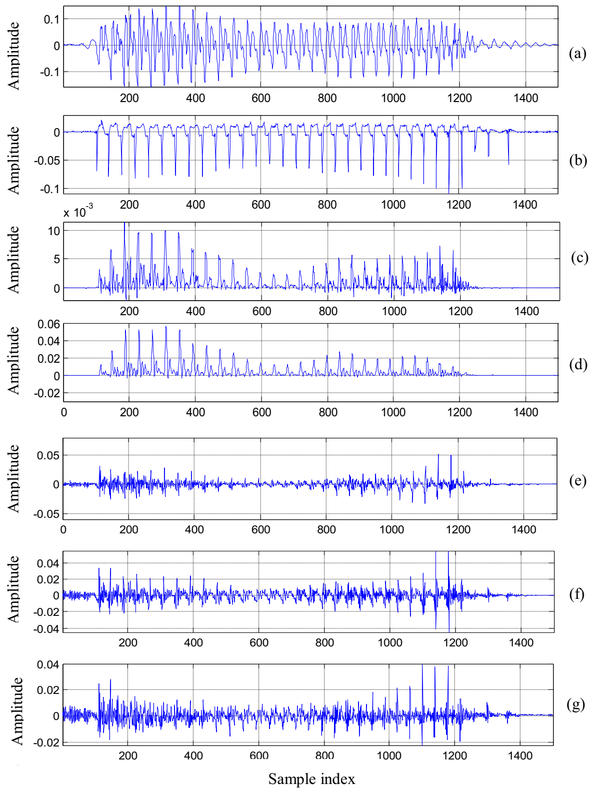

Figure 1: *(a) A speech segment of word 'philip' from CMU-ARCTIC database (arctic_a0001.wav) [22], Fs=8 kHz (b) the differenced EGG of (a) (i.e., ground truth for GCI), (c) TEO profile of (a), (d) TEO profile of (a) filtered at cutoff frequency 1 kHz by using a Gabor filter, (e) LP residual of (a) obtained by inverse filtering, (f) LP residual of (a) obtained by VWK series for d=1 and k=12 and (g) the NLP residual of (a) obtained by VWK series for d=2 and k=12.*

If the speech production mechanism (which can be modeled as *convolution* of excitation source with impulse response of vocal tract which itself is a cascade of several 2^{nd} order digital resonators each corresponding to a particular formant and having impulse response as damped sinusoid) would have been *linear*, then TEO profile would have been exponentially decaying signal [24]. The bumps in the TEO profile are significantly suppressed by filtering the signal through a Gabor filter (figure 1(d)) [4]. The LP residual shows high pulses at GCI as in figure 1(e)-(f). This is because LP residual (2) computed using the inverse filtering approach and by the proposed approach using VWK approximation gives high amplitude pulses at GCI that matches with the differenced EGG signal (i.e., ground truth). The *high energy pulses* in the TEO profile are present (due to *sudden* glottal closure at GCIs) at similar locations to that of the peaks obtained in the LP residual. It is evident from figure 1(f) and (g) that the LP residual is having larger non-zero samples (than its NLP counterpart) indicating that LP model is not sufficient to capture the complete dependencies in the samples of speech signal. The peaks in LP residual are sharper than in the TEO profile. Similar analysis is observed for the NLP residual (6) as in figure 1(g). It is observed that the NLP residual has have less energy than the LP residual and the peak amplitude around the GCI is less than that of LP residual. Thus, in NLP residual, the dependencies between the speech signals has been removed as a result the intelligibility of NLP residual is expected to be *relatively* less than that of the LP residual. To quantify this, we found the PESQ scores for LP and NLP residual as *4.4257* and *3.054*, respectively. The result suggests that NLP residual has less objective MOS (hence possibly giving less *intelligibility*) [25], [26].

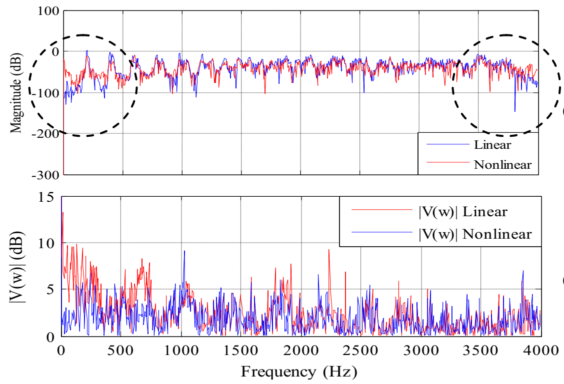

Figure 2: *(a) Spectra of LP and NLP residual (b) the difference magnitude |V(ω)| as a function of frequency.*

This relation can also be obtained in terms of the log-spectral distance (LSD) between the each of the *residual* spectra and *white* noise. Consider two spectra $E_i(\omega)$ and $W(\omega)$, then the difference between these two spectra on a log-magnitude *vs. f*requency scale, which is defined as [27],

$$V(\omega) = \log[E_i(\omega)] - \log[W(\omega)], \tag{7}$$

where $E_i(\omega)$ is the spectrum of LP residual when *i=LP* and NLP spectrum when *i=NLP*. $W(\omega)$ is the spectrum of white noise. One natural choice for a distortion or distance measure between $E_i(\omega)$ and $W(\omega)$ is that of a L_p norm, defined as,

$$d(S,S') = (d_p) = \left(\int_{-\pi}^{\pi} |V(\omega)|^p \frac{d\omega}{2\pi} \right)^{1/p}, \tag{8}$$

with *p=1*, the distortion is known as the *mean* log-spectral distortion and for *p=2*, it is called as *rms* log-spectral distortion [27]. Here, we estimate the LSD between the residuals and white noise as in figure 2(a). The log-spectral difference $|V(\omega)|$ computed using (7) is shown in figure 2(b). It is evident from figure 2(a) that spectrum of NLP residual is relatively *more flat* (especially in the frequency ranges *0-500* Hz and *3.5* kHz to *4* kHz as compared to LP spectrum (as indicated by dotted circles in figure 2(a)). In addition, $|V(\omega)|$ of NLP is below than that for LP at *almost all* the frequencies. The estimated values for the LSD for both LP and NLP methods are shown in Table 3. It can be observed that the spectral distance from white noise is relatively less for NLP residual than that of LP residual. This indicates that the spectrum of NLP residual is *relatively more* flat than its LP counterpart. Thus, the NLP residual is expected to have almost *no* relationship in the sequence of samples of residual signal. Thus, due to Wiener-Khinchin theorem, the NLP residual captures *higher-order statistical* relationships and *hidden* nonlinear dependencies in sequence of *samples* of speech samples and reduces the error in prediction.

Table 3. *The norm of LSD of LP and NLP residual spectra with white noise spectra*

Model Type	p=1 (L^1 norm)	p=2 (L^2 norm)
LP Approach	2.113	2.8500
NLP Approach	1.916	2.5867

4. Chaotic titration method

The use of VWK series is then used to estimate the chaotic nature of the speech signal and residual in prediction model is shown hereafter. The series acts here as an indication to chaos (just as *litmus paper* in the chemical titration process). After obtaining the coefficients of the model [21], the short-term prediction power of a model is measured by the standard deviation of one-step-ahead prediction error as,

$$\varepsilon^2(k,d) \equiv \frac{\sum_{n=1}^{N} (y_n^{calc}(k,d) - y(n))^2}{\sum_{n=1}^{N} (y(n) - \bar{y})^2}, \tag{9}$$

where $y_n^{calc} = \hat{y}(n)$, $\bar{y} = \frac{1}{N}\sum_{n=1}^{N} y_n$ and $\varepsilon^2(k,d)$ is normalized variance of the error or residuals. For each data series, the best model is $[k_{opt}, d_{opt}]$ which minimizes the following information criterion in accordance with the *parsimony principle* [28]:

$$C(r) = \log[\varepsilon(r)] + r / N, \tag{10}$$

where $r \in [1,M]$ is the number of polynomial terms of the truncated Volterra series expansions from a certain pair $\{k,d\}$. The procedure includes obtaining k^{lin} for the best *linear* model which minimizes $C(r)$ with d=1. Repeat it with increasing k and $d>1$ to obtain the best *nonlinear model*. Here, $d=1$ and $k=27$ is chosen for the linear model and for the nonlinear model $d=2$ and $k=6$ are selected. Therefore, each model will have *28* coefficients for prediction. Thereafter, the r-$C(r)$ curve is obtained for both the linear and nonlinear model from (9) and (10). For a speech segment from the CMU-ARCTIC database [22], the r-$C(r)$ curve is shown in Figure 3(b).

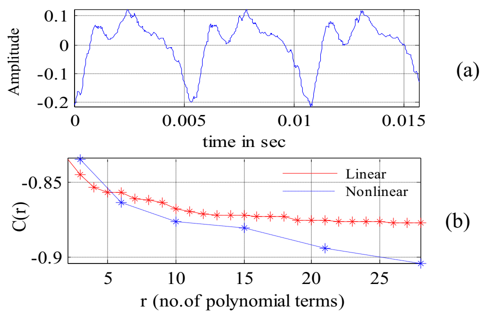

Figure 3: *a) Speech segment from CMU-ARCTIC database (arctic_a0002.wav)* [22] *and (b) r-C(r) plot of the speech segment in (a). (The stars indicate the memory terms.).*

It can be seen that there exists a difference between the linear and nonlinear curves. It is this difference that will estimate the amount of chaos in the signal. Chaos in the speech signal is estimated from the values of the *noise limit* (NL). NL value indicates as to when the information obtained from the linear and nonlinear predicted coefficients provides the same information. The VWK series indicates that the chaos in the signal is *neutralized* by the added noise when the curves of nonlinear and linear error values are nearly close to each other or in other words, the prediction done by the linear and nonlinear method is nearly the same. Thus, the neutral level and the value of the standard deviation (σ) of added noise at the neutral level is called as the *noise limit (NL)*.

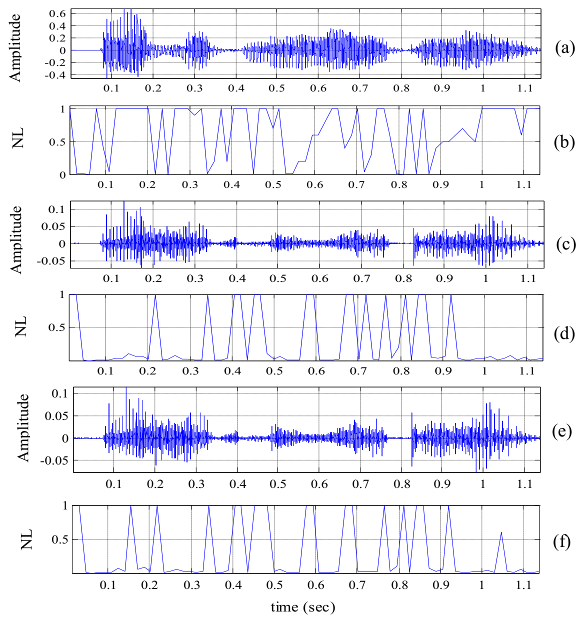

Figure 4: *(a) A speech segment 'I have no idea' from CMU-ARCTIC database (arctic_a0057.wav)* [22], *Fs=32 kHz (b) NL plot for (a), (c) the LP residual, (d) NL plot for (c), (e) the NLP residual and (f) NL plot for (e).*

To obtain the NL values for the given speech signal, we start with a frame, *titrate* the particular frame and find its NL value. According to the titration method, we add white noise of increasing σ to the signal till VWK nonlinear indicator cannot detect the nonlinearity. As in figure 4, we take a frame length of *30 ms* followed by a frame shift of *15 ms* and therefore we get a curve of NL *vs.* time. This shows that chaos exists not only in the LP residual [29], but also in the NLP residual.

5. Summary and Conclusions

In this paper, we have analyzed novel NLP-based on VWK approximation and shown improved prediction of speech signal than the traditional LP-based approach. It is observed that the speech signal and its residual also show chaotic characteristics. These chaotic characteristics can be used to analyze normal and pathological voices where the source (i.e., vocal fold motion) plays an important role [19]. In addition to the potential of proposed NLP approach for speech coding (as studied in [5], [12]), the authors believe that NLP may be explored in future for speech synthesis applications where the residual is replaced by noise or ideal pitch pulses to synthesize unvoiced and voiced speech signal, respectively [29]. Our future research work is directed towards the estimation of frequency response of vocal tract transfer function and modeling speech production mechanism by nonlinear prediction and synthesis of speech signal by knowing the chaos obtained by the titration approach using VWK series.

6. Acknowledgements

The authors would like to thank Department of Information and Technology (DIT), New Delhi, India and the authorities of DA-IICT, Gandhinagar for their kind help and cooperation to carry out this research work successfully.

7. References

[1] J. Markhoul, "Linear prediction: A tutorial review," in *Proc. of the IEEE*, vol. 63, no. 4, pp. 561-580, April 1975.

[2] B. Yegnanarayana and P. S. Murthy, "Enhancement of reverberant speech using LP residual signal," in *IEEE Trans. on Speech and Audio Process.*, vol. 8, no. 3, pp. 267-281, 2000.

[3] T. V. Ananthapadmanabha and G. Fant, "Calculation of true glottal flow and its components," *Speech Communication*, vol. 1, no. 3-4, pp. 167-184, 1982.

[4] T. F. Quatieri, *Discrete-Time Speech Signal Processing*. Prentice-Hall, 2002.

[5] J. Thyssen, H. Nielsen, and S. D. Hansen, "Non-linear short-term prediction in speech coding," in *Proc. of IEEE Int. Conf. Acoustics, Speech and Signal Process., ICASSP'94*, vol. 1, pp. 185-188, 1994.

[6] S. Martin, *The Volterra and Wiener Theories of nonlinear Systems*. John Wiley and Sons, 1980.

[7] R. H. Don and A. N. Venetsansopoulus, *Nonlinear Digital Filters: Principles and Application*. Kluwer Academic Publishers, 1990.

[8] V. Z. Marmarelis and X. Zhao, "Volterra models and three layers perception," in *IEEE Trans. on Neural Networks*, vol. 8, no. 6, pp. 1421-1433, 1997.

[9] N. Z. Hakim, J. J. Kaufman, G. Cerf, and H. E. Meadows, "Volterra characterization of neural networks," in *Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1128-1132, 1991.

[10] Y. Harkouss, J. Rousset, H. Chehade, E. Ngoya, and D. Barataud, "Modeling microwave devices and circuits for telecommunications systems design," in *Proc. of the International Joint Conference on Neural Networks*, vol. 1, pp. 128-133, 1998.

[11] G. Stegmayer, "Volterra series and neural networks to model an electronic device nonlinear behavior," in *Proc. of IEEE Conf. Neural Networks*, vol. 4, pp. 2907-2910, 2004.

[12] G. Alipoor and M. H. Savoji, "Speech Coding Using Non-linear Prediction Based on Volterra Series Expansion," in *13-th International Conference on Speech and Computer, SPECOM*, pp. 367-370, 2006.

[13] J. Pinto, G. S. V. S. Sivaram, H. Hermansky, and M. Magimai. -Doss, "Volterra Series for analyzing MLP based phoneme posterior probability Estimator," in *Proc. of IEEE Int. Conf. Acoustics, Speech and Signal Process., ICASSP'09*, pp. 1813-1816, 2009.

[14] M. Barahona and C.-S. Poon, "Detection of nonlinear dynalmics in short, noisy time series," *Nature*, vol. 381, pp. 215-217, May 1996.

[15] C.-S. Poon and M. Barahona, "Titration of chaos with added noise," in *Proceedings of National Academy of Science*, vol. 98, pp. 7107-7112, 2001.

[16] T. V. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction from linear prediction residual for identification of closed glottis interval," in *IEEE Trans. on Acoustic, Speech and Signal Process.*, vol. 27, no. 4, pp. 309-319, Aug. 1979.

[17] M. R. Schroder and B. S. Atal, "Code-excited linear prediction (CELP): High-quality speech at low rates", in *Proc. of IEEE Int. Conf. Acoustics, Speech and Signal Process., ICASSP'85*, vol. 10, pp. 937-940, 1985.

[18] M. Frechet, "Surles les fontionelles continues," *Annales Scientifiques de L'Ecole Normale Sup.*, vol. 27, pp. 193-216, 1910.

[19] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. John Wiley and Sons, 2000.

[20] H. A. Patil and T. B. Patel, "Novel chaotic titration method for analysis of normal and pathological voices," in *Int. Conf. on Signal Process. and Communication, SPCOM'12*, IISc, Banglore, India, pp. 1-5, 22-25 July, 2012.

[21] M. J. Korenberg, "Indentifying nonlinear difference equation and functional expansion representation: The fast orthogonal algorithm," *Annals of Biomedical Engineering*, vol. 16, pp. 123-142, 1988.

[22] "CMU-ARCTIC Speech Synthesis Database". Online: http://festvox.org/cmu_arctic/index.html (Last accessed May 28, 2013) .

[23] J. S. Garafolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database," in *National Institute of Standards and Technology (NIST), Gaithersburgh*, MD, USA, 1988

[24] H. A. Patil and K. K. Parhi, "Development of TEO phase for speaker recognition," in *Int. Conf. on Signal Process. and Communications, SPCOM'10*, IISc, Banglore, India, pp. 1-5, July 2010.

[25] A. W. Rix, J. G. Beerends, Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Proc. of IEEE Int. Conf. on Acoust., Speech and Signal Process., ICASSP'01*, vol. 2, pp. 749-752, 2001.

[26] "Perceptual Evaluation of Speech Quality (PESQ)". Online: http://www.itu.int/rec/T-REC-P.862/ (Last accessesed May 28, 2013).

[27] L. R. Rabiner, B.-H. Juang, and B. Yegnanarayana, *Fundamentals of Speech Recognition*. Prentice Hall, 2010.

[28] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. on Automatic Control*, vol. 19, no. 6, pp. 716-723, 1974.

[29] C. Tao, J. Mu, and G. Du, "Chaotic characteristics of speech signal and its LPC residual," *Acoustical Letter in Acoustical Science and Technology*, vol. 25, no. 1, pp. 50-53, 2004.

.