6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

ADAPTIVE NONLINEAR PREDICTION BASED ON ORDER STATISTICS FOR SPEECH SIGNALS

Tetsuya Shimamura and Haruko Hayakawa

Department of Information and Computer Sciences
Saitama University, 255 Shimo-Okubo, Urawa, 338-8570, Japan

10.21437/Eurospeech.1999-90

ABSTRACT

This paper proposes a novel adaptive algorithm for nonlinear prediction of speech signals, which turns out to be the adaptation procedure for an order statistic LMS predictor. The LMS-L filter Pitas et al. addressed is modified to preserve the time information in the input vector for the adaptation, in which a coefficient matrix is utilized to update the predictor coefficients. Computer simulations demonstrate that the novel nonlinear predictor provides better performance than the Volterra quadratic predictor as well as the linear predictor.

1. INTRODUCTION

Speech production is extensively assumed to be modeled by the use of a linear filter. In fact, the technique of linear prediction (LP)[1] has been used in many speech processing systems, in which the speech signal is modeled as the output of a linear all-pole filter whose input is white noise for unvoiced speech or a chain of impulses for voiced speech. In most of cases, the LP method has made a successful result. However, the LP method obviously has the remaining part to be improved, because for voiced speech analysis, the performance of the LP method is affected by the pitch period of the speech signal[2][3]. If the LP method is implemented in the standard form of short-term prediction, the LP method may not remove the impulsive residual errors invoked by a link with the excitation source being the impulse chains. The reason for this may be that the LP method has been derived based on the autoregressive process whose input is white noise.

Thyssen et al.[8] unveiled that the impulses to occur synchronously with the pitch of voiced speech invoke a kind of nonlinearity in the speech signal, and that the nonlinearity should be removed by a nonlinear prediction method. In [8], the Volterra filter and multi-layer perceptron were deployed to remove the nonlinear dependency of the speech signal. Diaz-de-Maria et al.[9] proposed the use of radial basis functions for the same purpose. However, the neural network approach such as the multi-layer perceptron and radial basis functions has a common problem associated with the increased computational complexity. Chu et al.[10] showed that the performance of the neural network predictor is apt to be improved, as the complexity of the analysis system is increased.

In this paper, an adaptive implementation for the nonlinear prediction of speech signals is investigated. Although the LP method is popular in the form of batch processing, its sequential form, adaptive linear predictor, is also useful for speech analysis[4]. It has an important application in the adaptive differential pulse-code modulation system[5]. If the least mean squares (LMS) algorithm is deployed for the adaptation procedure, the computation of the predictor is very simple. Although the adaptive linear predictor has the above-mentioned nonlinearity problem for voiced speech analysis, it may be avoided by a nonlinear predictor such as the Volterra predictor[7]. In this paper, we present a novel LMS-type predictor to accomplish the nonlinear prediction of voiced speech signals, into the adaptation procedure of which, the operation of order statistics is incorporated in the form where the time information in the input vector is preserved. The basic idea of the novel predictor is that the order statistic operations have the potential to reduce impulsive noise, which is a natural appearance in the residual of the linear predictor. By computer simulations, we show how the order statistic predictor produces the residual errors in a comparative form with the conventional predictors.

2. LINEAR AND NONLINEAR PREDICTORS

At first, let us consider an adaptive predictor as depicted in Figure 1, in which the speech signal $s(n)$ is assumed to be predicted from its previous values such as

$$\hat{s}(n) = \phi(s(n-1), s(n-2), ..., s(n-M)) \tag{1}$$

where the hat denotes an estimate and $\phi(\cdot)$ means a mapping function including adjustable parameters. If the mapping function $\phi(\cdot)$ transforms nonlinearly as well as linearly the input sequence, the predictor in Figure 1 may be recognized as a general predictor.

2.1. LMS Predictor

A special case of the general predictor is the adaptive linear predictor, in which the estimate of the speech signal is given by

$$\hat{s}(n) = \sum_{i=1}^{M} a_i s(n-i) \tag{2}$$

where $a_i, i = 1, 2, ..., M$ correspond to the prediction coefficients. If the LMS algorithm (normalized version)is used for the adaptation, then the coefficients for the predictor is updated as follows.

$$e(n) = s(n) - \mathbf{s}(n)^T \mathbf{a}(n) \tag{3}$$

$$\mathbf{a}(n+1) = \mathbf{a}(n) + \frac{\mu}{\mathbf{s}(n)^T \mathbf{s}(n) + \beta} \mathbf{s}(n) e(n) \tag{4}$$

where T denotes transposition, $\mathbf{s}(n)$ is the input vector

$$\mathbf{s}(n) = [s(n-1), s(n-2), ..., s(n-M)]^T \tag{5}$$

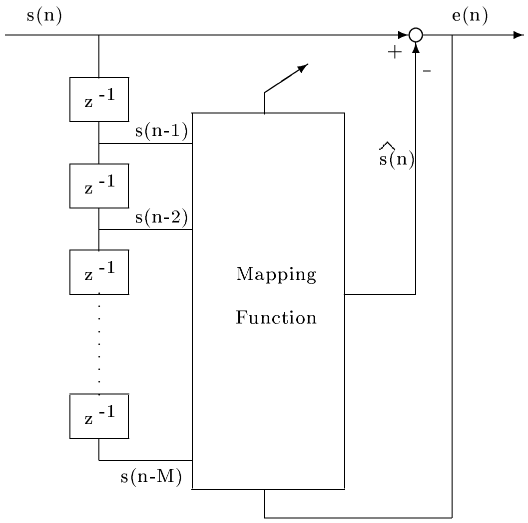

Figure 1: Configuration of the general adaptive predictor.

and $\mathbf{a}(n)$ is the coefficient vector for n iterations

$$\mathbf{a}(n) = [a_1(n), a_2(n), ..., a_M(n)]^T. \quad (6)$$

The μ and β mean the step size and stabilized parameters for the normalized LMS algorithm, respectively.

2.2. Volterra Predictor

When the function $\phi(\cdot)$ in (1) behaves nonlinearly, an adaptive nonlinear predictor can be realized. Mumolo et al.[7] deployed the Volterra quadratic filter for the function $\phi(\cdot)$. For the Volterra predictor, the estimate of the speech signal is given by

$$\hat{s}(n) = \sum_{i=1}^{M_1} a_i s(n-i) + \sum_{i=1}^{M_2} \sum_{j=i}^{M_2} b_{ij} s(n-i) s(n-j) \quad (7)$$

where $a_i, i = 1, 2, ..., M_1$ and $b_{ij}, i, j = 1, 2, ..., M_2$ correspond to the linear and quadratic prediction coefficients, respectively. The adaptation procedure for the Volterra predictor is given by

$$e(n) = s(n) - \mathbf{q}(n)^T \mathbf{h}(n) \quad (8)$$

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \frac{\mu}{\mathbf{q}(n)^T \mathbf{q}(n) + \beta} \mathbf{q}(n) e(n) \quad (9)$$

where $\mathbf{q}(n)$ is the input vector

$$\mathbf{q}(n) = [s(n-1), s(n-2), ..., s(n-M_1), \\ s(n-1)^2, s(n-1)s(n-2), ..., s(n-M_2)^2]^T \quad (10)$$

and $\mathbf{h}(n)$ is the coefficient vector for n iterations

$$\mathbf{h}(n) = [a_1(n), a_2(n), ..., a_{M_1}(n), b_{11}, ..., b_{M_2 M_2}]^T. \quad (11)$$

3. ORDER STATISTIC PREDICTOR

Pitas et al.[6] proposed the LMS-L filter for the purpose of impulsive noise reduction, in which the order statistic of the input vector is used as the input vector for the LMS adaptation. The LMS-L filter efficiently and effectively reduces additive impulse noise, but has a drawback that it cannot preserve the time information being the order of the elements of the input vector for the adaptation. This property for the LMS-L filter may not suitable for speech analysis, because the assumption of speech prediction is based on (1) where the speech signal is predicted from its previous values. This means that the oder of the elements of the input vector should be preserved in the adaptation. Based on such an idea, we derive a novel order statistic adaptive predictor. The adaptive predictor has a matrix based filter structure, which involves a nonlinear adaptation operation, but the coefficients to be updated are linear for the predictor output and the number of the coefficients to be updated is the same as that of the linear predictor. Therefore, the adaptive predictor would be recognized as a linear predictor whose coefficient adaptation is done by a nonlinear algorithm. In this paper, we refer the nonlinear adaptive algorithm to as the OSLMS algorithm.

The OSLMS algorithm prepares a coefficient matrix instead of a coefficient vector as follows.

$$\mathbf{C}(n) = \begin{bmatrix} c_{11}(n) & c_{12}(n) & ... & c_{1,M}(n) \\ c_{21}(n) & c_{22}(n) & ... & c_{2,M}(n) \\ \vdots & \vdots & \ddots & \vdots \\ c_{M,1}(n) & c_{M,2}(n) & ... & c_{M,M}(n) \end{bmatrix} \quad (12)$$

The elements $c_{ij}(n), i, j = 1, 2, ..., M$ are initialized to zeros at $n = 0$.

For the adaptation, among the $M \times M$ elements of $\mathbf{C}(n)$, only M elements are selected and updated. Specifically, for n iterations, only $c_{m(j)j}(n), j = 1, 2, ..., M$ are selected. And a coefficient vector

$$\mathbf{c}(n) = [c_{m(1)1}(n), c_{m(2)2}(n), ..., c_{m(M)M}(n)]^T \quad (13)$$

is made and updated where $m(j)$ corresponds to the order when the input vector

$$\mathbf{s}(n) = [s(n-1), s(n-2), ..., s(n-M)]^T \quad (14)$$

is transformed into the order statistic vector

$$\mathbf{x}(n) = [x_1(n), x_2(n), ..., x_M(n)]^T \quad (15)$$

$$x_1(n) \le x_2(n) \le ... \le x_M(n). \quad (16)$$

The $m(j)$ is determined for $i, j = 1, 2, ..., M$ by

$$m(j) = i \quad \text{if } \ x_i(n) = s(n-j). \quad (17)$$

The adaptation equation of the OSLMS algorithm is given by

$$e(n) = s(n) - \mathbf{x}(n)^T \mathbf{c}(n) \quad (18)$$

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \frac{\mu}{\mathbf{s}(n)^T \mathbf{s}(n) + \beta} \mathbf{x}(n) e(n). \quad (19)$$

The updated coefficient vector $\mathbf{c}(n+1)$ is inserted into the coefficient matrix $\mathbf{C}(n+1)$. By this, one iteration of the OSLMS algorithm terminates. All the elements of the coefficient matrix would be updated as the adaptation is progressed.

4. SIMULATION RESULTS

Computer simulations were carried out to verify the performance of the OSLMS predictor. A synthetic vowel /o/ and a real vowel /a/ pronounced by a male were used.

The synthetic speech was generated based on the following equations.

$$u(n) = \sum_{j=0}^{\infty} \delta(n - jQ) \qquad n = 0, 1, 2, ... \tag{20}$$

$$s(n) = \sum_{k=1}^{M} a_k s(n-k) + Gu(n) \tag{21}$$

where Q corresponds to the samples for the pitch period, $a_k, k = 1, 2, ..., M$ and G are constant parameters. The specific values are shown below. $G = 0.1345$, $a_1 = 1.53527$, $a_2 = -0.97789$, $a_3 = 1.48396$, $a_4 = -1.78023$, $a_5 = 0.71704$, $a_6 = 0.73514$, $a_7 = 0.76348$, $a_8 = 0.12135$, $a_9 = -0.15552$, $a_{10} = -0.178143$, sampling frequency f_s = 10 kHz, pitch period = 2 msec ($Q = 20$). The synthetic vowel was used in [11] with a different pitch period. The real vowel /a/ was also sampled with f_s =10 kHz. The band limitation was 3.4.kHz.

Figure 2 shows the convergence on the synthetic vowel /o/ for the LMS predictor in which the order is $M = 10$, the step size $\mu = 1.0$ and the stabilized parameter $\beta = 0.05$. It is observed that the convergence speed is fast (only 200 iterations are required for convergence), but there exist impulsive residual errors in the steady state.

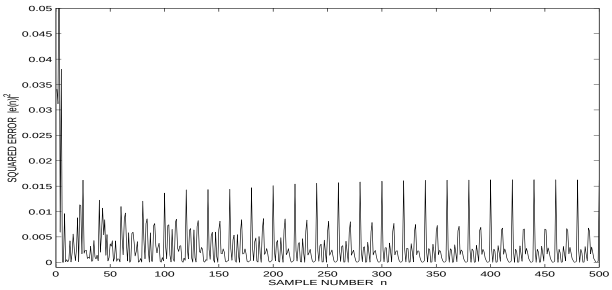

Figure 2: Convergence for the LMS predictor on the synthetic vowel.

Figure 3 shows the convergence for the Volterra predictor in which the order for the linear predictor part is $M_1 = 10$ and that for the quadratic counterpart is $M_2 = 3$. The adaptation is again with the step size $\mu = 1.0$ and with the stabilized parameter $\beta = 0.05$. The Volterra predictor requires much more iterations to reach the steady state. In Figure 3, it is observed that 600 iterations are required for convergence. The amplitude of the residual errors is, however, reduced.

Figure 4 shows the convergence for the OSLMS predictor in which the order is $M = 10$, the step size $\mu = 1.0$ and the stabilized parameter $\beta = 0.05$. The convergence speed is faster than the Volterra predictor, but slightly slower than the LMS predictor. It is obviously observed that the residual errors are negligible. In the steady state, we evaluated the following prediction gain

$$PG = 10\log_{10} \frac{\sum_{n=n_1}^{n_2} s(n)^2}{\sum_{n=n_1}^{n_2} e(n)^2} \tag{22}$$

for each predictor where n_1 and n_2 were adjusted so that each predictor keeps its steady state. The gain was PG = 9.6 dB for the LMS predictor, PG = 12.4 dB for the

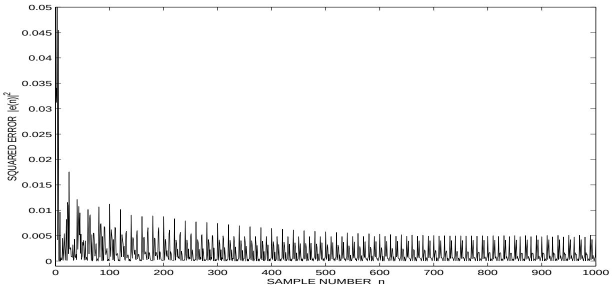

Figure 3: Convergence for the Volterra predictor on the synthetic vowel.

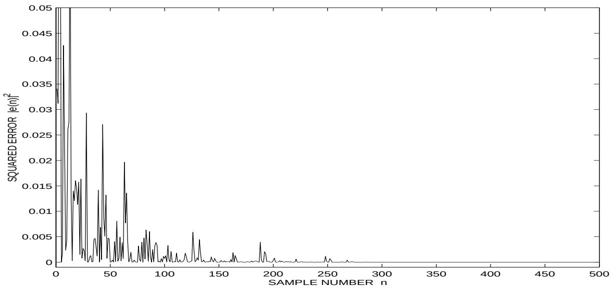

Figure 4: Convergence for the OSLMS predictor on the synthetic vowel.

Volterra predictor and PG = 124.6 dB for the OSLMS predictor, respectively.

The estimated speech wave is compared with the original one. Figure 5 shows the original wave of the synthetic vowel and the wave evaluated by the LMS predictor. Figure 6 corresponds to the case of the OSLMS predictor. Obviously the LMS predictor contains a significant error, while the OSLMS predictor invokes a mostly matched wave estimate.

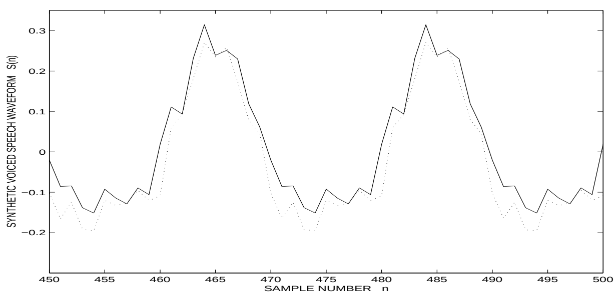

Figure 5: Original speech wave (solid line) and speech wave estimated by the LMS predictor (dotted line).

The performance on the real speech for the LMS predictor with the order $M = 14$ is unveiled in Figure 7. On the the hand, the performance for the OSLMS predictor with the same order is shown in Figure 8. We see that the OSLMS predictor significantly reduces the residual errors and provides better performance than the LMS

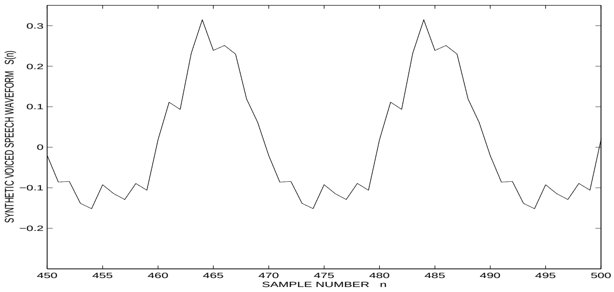

Figure 6: Original speech wave (solid line) and speech wave estimated by the OSLMS predictor (dotted line).

predictor.

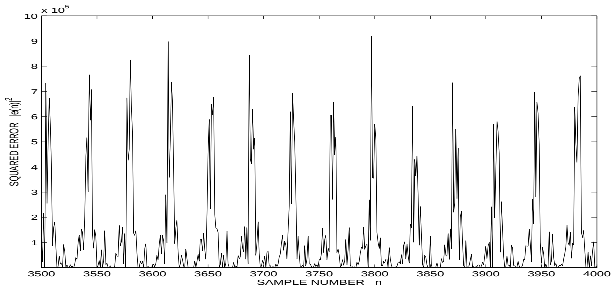

Figure 7: Residual errors for the LMS predictor on the real speech /a/.

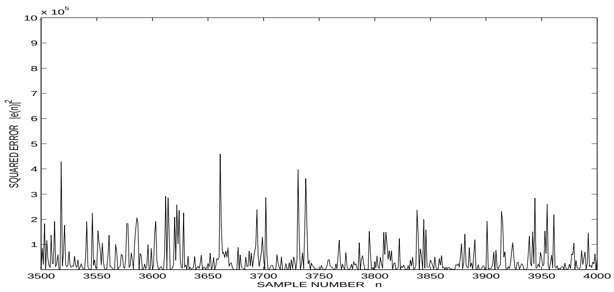

Figure 8: Residual errors for the OSLMS predictor on the real speech /a/.

5. CONCLUSION

A novel adaptive algorithm to accomplish nonlinear prediction has been presented in which the order statistic operation with the LMS adaptation is implemented. Computer simulations have demonstrated that with moderate convergence speech, the novel predictor provides superior prediction gain to the Volterra quadratic predictor as well as the LMS predictor.

6. REFERENCES

[1] B.S.Atal and S.Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave", J. Acoustics Society of America, Vol.50, No.2, pp.637-655, 1971.

[2] S.Chandra and W.C.Lin, "Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis", IEEE Trans. Vol.ASSP-22, No.6, pp.403-415, 1974.

[3] L.R.Rabiner, B.S.Atal and M.R.Sambur, "LPC prediction error - analysis of its variation with the position of the analysis frame", IEEE Trans. Vol.ASSP-25, No.5, pp.434-442, 1977.

[4] J.D.Gibson, J.L.Melsa and S.K.Jones, "Digital speech analysis using sequential estimation techniques", IEEE Trans. Vol.ASSP-23, No.4, pp.362-369, 1975.

[5] J.D.Gibson, S.K.Jones and J.L.Melsa, "Sequentially adaptive prediction and coding of speech signals", IEEE Trans. COM-22, No.11, pp.1789-1797, 1974.

[6] I.Pitas and A.N.Venetsanopoulos, "Adaptive filters based on order statistics", IEEE Trans. Vol.ASSP-39, No.2, pp.518-522, 1991.

[7] E.Mumolo, A.Carini and D.Francescato, "ADPCM with non linear predictors", Proc. EUSIPCO94, pp.387-390, 1994.

[8] J.Thyssen, H.Nielsen and S.D.Hansen, "Non-linear short-term prediction in speech coding", Proc. ICASSP94, pp.I-185-188, 1994.

[9] F.Diaz-de-Maria and A.R.Figuiras-Vidal, "Nonlinear prediction for speech coding using radial basis functions", Proc. ICASSP95, pp.788-791, 1995.

[10] W.C.Chu and N.K.Bose, "Speech signal prediction using feedforward neural network", Electronics Letters, Vol.34, No.10, pp.999-1001, 1998.

[11] K.K.Paliwal and P.V.S.Rao, "A modified autocorrelation method of linear prediction for pitch-synchronous analysis of voiced speech", Signal Processing, Vol.3, No.2, pp.181-185, April, 1981.