Low-Order Volterra Long-Term Predictors

Vladimir Despotovic[*], *Norbert Görtz*[†], *Zoran Peric*[‡]

[*] University of Belgrade, Technical Faculty in Bor, 19210 Bor, Serbia
Email: vdespotovic@tf.bor.ac.rs
[†] Vienna University of Technology, Institute of Telecommunications, 1040 Vienna, Austria
Email: norbert.goertz@nt.tuwien.ac.at
[‡] University of Nis, Faculty of Electronic Engineering, 18000 Nis, Serbia
Email: zoran.peric@elfak.ni.ac.rs

Abstract

Models based on linear prediction have been used for several decades in different areas of speech signal processing. While the linear approach has led to great advances in the last 40 years, it neglects nonlinearities present in the speech production mechanism. This paper compares the results of long-term nonlinear prediction based on second-order and third-order Volterra filters. Additional improvement can be obtained using fractional-delay long-term prediction. Experimental results reveal that the proposed method outperforms linear long-term prediction techniques in terms of prediction gain.

1 Introduction

Linear filters have played a crucial role in the development of various signal processing techniques. Their obvious advantage is the inherent simplicity from both a conceptual and an implementation point-of-view, which leads to satisfactory performance in a number of applications including speech modeling and prediction [1]. However, for large classes of problems nonlinear models have shown to be more appropriate.

The Volterra series model is most widely used in nonlinear adaptive filtering. The Volterra series expansion can be regarded as a generalized Taylor series of a function with memory. A discrete-time Volterra series with infinite memory has the expression

$$y(n) = h_0 + \sum_{k=0}^{\infty} h_1(k)x(n-k) + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} h_2(i,j)x(n-i)x(n-j) + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} h_3(i,j,k)x(n-i)x(n-j)x(n-k) + \ldots \qquad (1)$$

where $x(n)$ is the input signal, $y(n)$ the output of the model, h_0 the bias coefficient, h_1 the linear coefficients, h_2 the quadratic coefficients, h_3 the cubic coefficients etc.

The huge number of coefficients of Volterra filters with long memory inhibits their practical use in signal processing [2]. That is the main reason why they are truncated to low nonlinearity orders (usually 2nd or 3rd orders).

2 Long-Term Prediction of Speech Based on Volterra Filters

In low bit-rate coders, the near-sample and far-sample redundancies of the speech signal are usually removed by a cascade of a short-term and a long-term linear predictor. The short-term predictor (STP) in such a configuration will first remove the redundancies due to the formants, while the long-term predictor (LTP) will subsequently remove the redundancies due to the presence of a pitch excitation [3]. This solution is used in a number of speech coders (CELP, RPE-LTP etc.).

A number of short-term nonlinear speech predictors based on Volterra series are reported in the literature [1], [4,5], showing potential benefits over linear predictors. However, the exponential increase in the number of coefficients prevents their wider usage in speech coding. On the other hand, long-term predictors are usually used as one-tap predictors; that is, prediction is based on one single sample from the distant past. This gives a possibility of a design of nonlinear long-term predictors based on Volterra filters without a substantial increase in the number of coefficients [6].

2.1 Second-Order Long-Term Prediction

In this paper we substitute the linear long-term predictor by a nonlinear one, based on a second-order Volterra filter and connect it in cascade with a short-term linear predictor, as shown in Fig. 1. It predicts the current signal sample from a past sample that is one or more pitch periods apart. The corresponding predicted sample equals

$$\hat{e}_x(n) = h_1 \cdot e_x(n-T) + h_2 \cdot e_x{}^2(n-T)\,, \qquad (2)$$

where e_x is the residual signal (prediction error) after short-term linear prediction, T is the pitch period and h_1 and h_2 are LTP coefficients. Compared to linear LTP, the number of coefficients is increased only by one (h_2). Within a given time interval of interest, we seek to find h_1 and h_2 such that the sum $J = \sum_n \left(e_x(n) - \hat{e}_x(n)\right)^2$ of squared errors is minimized. Substituting (2), differentiating with respect to h_1 and h_2 and equating to zero, the following LTP coefficients are obtained [6]:

$$h_1 = \frac{\sum_n e_x{}^4(n-T)\cdot\sum_n e_x(n)\cdot e_x(n-T) - \sum_n e_x{}^3(n-T)\cdot\sum_n e_x(n)\cdot e_x{}^2(n-T)}{\sum_n e_x{}^4(n-T)\cdot\sum_n e_x{}^2(n-T) - \left(\sum_n e_x{}^3(n-T)\right)^2} \qquad (3)$$

$$h_2 = \frac{\sum_n e_x{}^2(n-T)\cdot\sum_n e_x(n)\cdot e_x{}^2(n-T) - \sum_n e_x{}^3(n-T)\cdot\sum_n e_x(n)\cdot e_x(n-T)}{\sum_n e_x{}^4(n-T)\cdot\sum_n e_x{}^2(n-T) - \left(\sum_n e_x{}^3(n-T)\right)^2} \qquad (4)$$

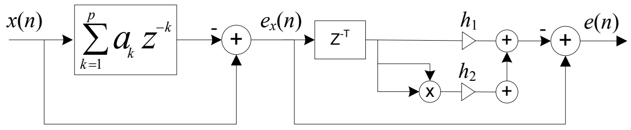

Figure 1: Short-term linear predictor connected in cascade with a long-term second-order Volterra predictor

2.2 Third-Order Long-Term Prediction

Since the number of coefficients is not a critical issue in Volterra long-term prediction, the third-order predictor can be used as well in cascade with a short-term linear predictor, as shown in Fig. 2. In this case the number of coefficients is increased only by two compared to linear LTP. The corresponding predicted sample equals

$$\hat{e}_x(n) = h_1 \cdot e_x(n-T) + h_2 \cdot e_x{}^2(n-T) + h_3 \cdot e_x{}^3(n-T) \quad (5)$$

Minimizing the sum of squared errors the following LTP coefficients are obtained

$$h_1 = \frac{q_2 \cdot q_4 \cdot q_7 - q_3 \cdot q_4 \cdot q_5 + q_3^2 \cdot q_6 - q_3 \cdot q_7 \cdot q_8 - q_2 \cdot q_5 \cdot q_6 + q_5^2 \cdot q_8}{q_3^3 + q_2^2 \cdot q_7 + q_1 \cdot q_5^2 - q_1 \cdot q_3 \cdot q_7 - 2 \cdot q_2 \cdot q_3 \cdot q_5} \quad (6)$$

$$h_2 = \frac{q_1 \cdot q_5 \cdot q_6 - q_3 \cdot q_5 \cdot q_8 - q_2 \cdot q_3 \cdot q_6 + q_2 \cdot q_7 \cdot q_8 - q_1 \cdot q_4 \cdot q_7 + q_3^2 \cdot q_4}{q_3^3 + q_2^2 \cdot q_7 + q_1 \cdot q_5^2 - q_1 \cdot q_3 \cdot q_7 - 2 \cdot q_2 \cdot q_3 \cdot q_5} \quad (7)$$

$$h_3 = \frac{q_1 \cdot q_4 \cdot q_5 - q_2 \cdot q_3 \cdot q_4 + q_2^2 \cdot q_6 - q_2 \cdot q_5 \cdot q_8 - q_1 \cdot q_3 \cdot q_6 + q_3^2 \cdot q_8}{q_3^3 + q_2^2 \cdot q_7 + q_1 \cdot q_5^2 - q_1 \cdot q_3 \cdot q_7 - 2 \cdot q_2 \cdot q_3 \cdot q_5} \quad (8)$$

where

$$q_1 = \sum_n e_x{}^2(n-T) \quad (9)$$

$$q_2 = \sum_n e_x{}^3(n-T) \quad (10)$$

$$q_3 = \sum_n e_x{}^4(n-T) \quad (11)$$

$$q_4 = \sum_n e_x(n) \cdot e_x{}^2(n-T) \quad (12)$$

$$q_5 = \sum_n e_x{}^5(n-T) \quad (13)$$

$$q_6 = \sum_n e_x(n) \cdot e_x{}^3(n-T) \quad (14)$$

$$q_7 = \sum_n e_x{}^6(n-T) \quad (15)$$

$$q_8 = \sum_n e_x(n) \cdot e_x(n-T) \quad (16)$$

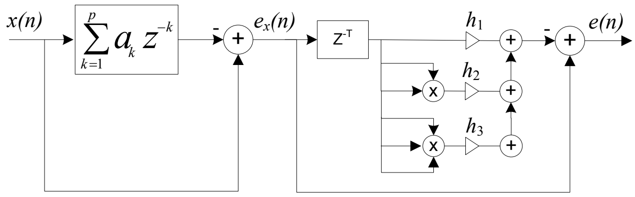

Figure 2: Short-term linear predictor connected in cascade with a long-term third-order Volterra predictor

3 Fractional-Delay Long-Term Prediction

One source of error that limits the resolution, and hence the accuracy of the long-term predictor, is time discretization of the pitch estimates, introduced by sampling the speech signal asynchronously with the instantaneous pitch value. A pitch period, expressed as an integer multiple of the sampling interval, contains a time discretization error which may lead to audible distortions in speech coding applications [7]. In fact, the pitch period of the original continuous-time signal (before sampling) is a real number; thus, integer periods are only approximations introducing errors that might have negative impact on the system performance.

Fractional pitch periods are introduced in long-term linear prediction as a means to increase the temporal resolution. In this case, the pitch period is allowed to have a fractional part besides the integer part [8]. The same principle can be used, without any restriction, for non-linear Volterra long-term prediction.

Fractional-delay long-term prediction can be obtained via interpolation. If *n* fractional samples are to be computed between two integer samples, this is equivalent to increasing the sampling frequency by *n*. Experimental results have shown that by using fractional pitch periods, the average prediction gain can be increased by up to 2.5 dB. The improvement is higher for female speech since the shorter pitch period makes it more susceptible to discretization errors [9].

Interpolation can be performed in various ways. For instance, an interpolation function as a product of Hamming window and a *sinc* function is used in the Federal Standard CELP 1016 [10]. Medan, Yair, and Chazan use a simple linear interpolation technique [7]. The following low-pass interpolation algorithm is used in this paper:

- The input speech vector is expanded to the correct length by inserting zeros between the original data values.
- A special symmetric FIR filter is designed that allows the original data to pass through unchanged, and interpolates in between, so that the mean-square errors between the interpolated points and their ideal values are minimized.
- The filter is applied to the input vector to produce the interpolated output vector.

Since the interpolation results in an upsampled signal, it has to be downsampled to the actual sample rate during the reconstruction process, which is simply achieved by using every *n*-th sample only (no filtering involved).

4 Results and Discussion

Let us assume a short-term linear predictor with the order *p*=10 connected in cascade with the one-tap long-term nonlinear predictor based on the second-order and the third-order Volterra filters, as shown in Fig. 1 and Fig. 2 respectively. The predictor operates on frames divided into four subframes of equal length.

Fig. 3 shows prediction error for nonlinear second-order and third-order Volterra long-term predictors connected in cascade with a linear short-term predictor with prediction order $p = 10$ on one characteristic frame of speech. Long-term prediction gains were found to be 3.50 and 4.21 *dBs* for the second-order and the third-order predictors respectively. Increasing the nonlinearity order from two to three leads approximately to an increase of 0.7 *dB* in prediction gain in this case. Experiments using a larger amount of speech samples (225 seconds of speech extracted from the TIMIT database, 70 sentences spoken by 44 male and 26 female American English speakers) gave similar results. The results for 20 *ms* and 30 *ms* frame lengths are summarized in Tab. 1. Compared to linear LTP, up to 0.8 *dB* higher prediction gain was reported using nonlinear second-order Volterra long-term predictor. Additional 0.5 *dB* can be obtained using the third-order model. However, one should be careful with increasing the model's order, since the use of higher-order models increases the computational requirements as well.

The results shown in Tab. 1 assume integer-valued pitch periods. A possible problem with integer pitch period is the phenomenon of pitch multiplication. For periodic signals the current period is not only similar to the previous one, but also to periods that occurred multiple periods ago. Pitch multiplication is disadvantageous for coding since a sudden jump of pitch might lead to artifacts in the synthesized speech [8]. This effect is observed in the first column of the Tab. 2, where pitch periods are given on one characteristic frame of speech. The pitch period in the third subframe is clearly a multiple of a true value of the pitch. This problem can be overcome using fractional-delay LTP, as explained in Section 3. When fractional-delays are introduced with interpolation factor n=2, the problem with pitch multiplication will not occur so often anymore, as shown in an example in the second column of the Tab. 2. The better match of the true pitch period obviously leads to a smaller prediction error (Fig. 4), thus higher prediction gain, for both second-order and third-order nonlinear Volterra long-term predictors.

Tab. 3 shows prediction gains of the fractional-delay second-order and third-order Volterra pitch predictors obtained on the same speech database as in Tab. 1. Fractional-delay pitch prediction with interpolation factor n=2 increases LTP gain up to 0.9 *dB*. Our experiments have shown that further increase of the interpolation factor yields only a modest additional increase in prediction gain (not greater than 0.2 *dB*).

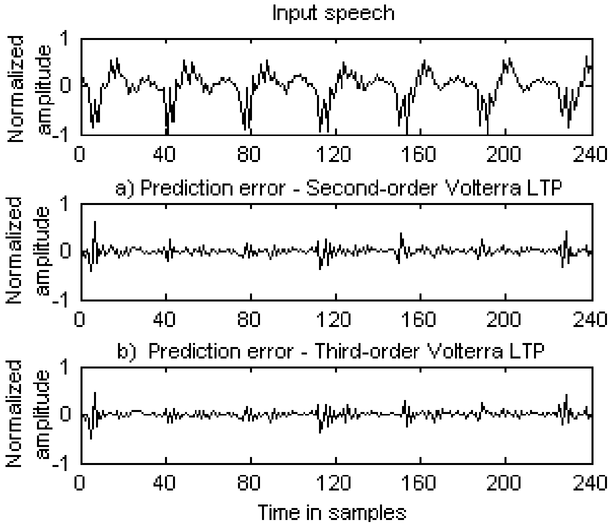

Figure 3: Prediction error for a linear short-term predictor connected in cascade with a nonlinear Volterra long-term predictor a) second-order model and b) third-order model

	Frame length	
	20 *ms*	30 *ms*
Linear LTP gain [*dB*]	3.1	2.7
2^{nd} order Volterra LTP gain [*dB*]	3.9	3.4
3^{rd} order Volterra LTP gain [*dB*]	4.4	3.7

Table 1: Long-term prediction gain for linear predictors and non-linear second-order and third-order Volterra predictors obtained on 225 seconds of speech

Subframe	Pitch period without fractional delay	Pitch period with fractional delay
1	35	35
2	36	36
3	74	37.5
4	38	38

Table 2: Pitch period with and without fractional delay

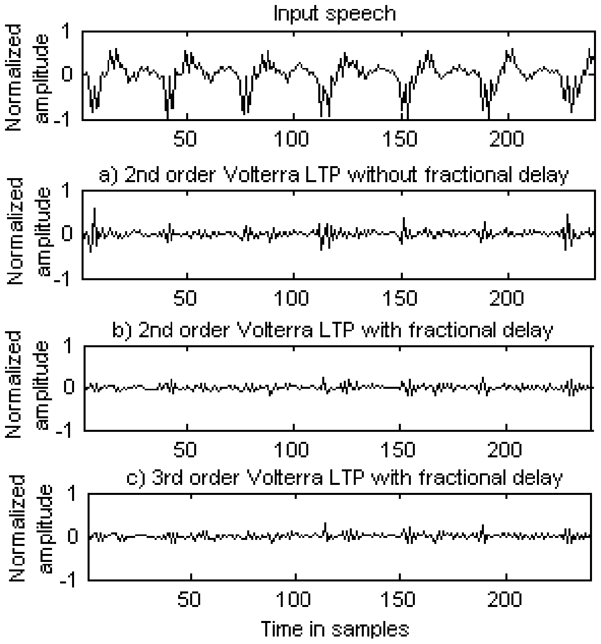

Figure 4: Prediction error for a linear short-term predictor connected in cascade with a nonlinear Volterra long-term predictor a) second-order model without fractional delay, b) second-order model with fractional delay and c) third-order model with fractional delay

	Frame length	
	20 *ms*	30 *ms*
2^{nd} order Volterra LTP gain with fractional delay [*dB*]	4.8	4.1
3^{rd} order Volterra LTP gain with fractional delay [*dB*]	5.1	4.4

Table 3: Long-term prediction gain for non-linear second-order and third-order Volterra predictors with fractional-delays (interpolation factor n=2) obtained on 225 seconds of speech

One can argue that introducing the fractional-delay long-term predictor has a greater impact on the quality of the reconstructed speech than increasing the nonlinearity order of the model, but still combining these two techniques will give significantly better overall results.

5 Conclusions

Closed-form expressions for the optimal predictor coefficients of second-order and third-order long-term Volterra predictors are derived in this paper. They can be used in a cascade with short-term linear predictors, with only a minimal increase in the number of coefficients. An increase in prediction gain of 0.5 *dB* was achieved using the third-order predictor compared to the second-order one. However, this increases computational complexity as well, but only slightly so.

The prediction gain can be additionally improved using fractional-delay long-term prediction. Our experiments have shown 0.9 *dB* additional prediction gain

References

[1] Gh. Alipoor and M. H. Savoji, "Employing Volterra filters in the ADPCM technique for speech coding: a comprehensive investigation," *European Transactions on Telecommunications*, vol. 22, no. 2, pp. 81-92, Mar. 2011.

[2] A. Stenger and R. Rabenstein, "Adaptive Volterra filters for nonlinear acoustic echo cancellation, " in *Proc. of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Antalya, Turkey, pp. 679-683, Jun. 1999.

[3] D. Giacobello, M. G. Christensen, J. Dahl, S. H. Jensen and M. Moonen, "Joint estimation of short-term and long-term predictors in speech coders," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP '09)*, Taipei, Taiwan, pp. 4109-4112, Apr. 2009.

[4] J. Thyssen, H. Nielsen and S. Hansen, "Non-linear short-term prediction in speech coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP '94)*, Adelaide, Australia, pp. 185-188, Apr. 1994.

[5] E. Mumolo, A. Carini and D. Francescato, "ADPCM with nonlinear predictors," *in Signal Processing VII: Theories and applications*, Ed. Elsevier, vol. 1, pp. 387-390, Sep. 1994.

[6] V. Despotovic, N. Goertz and Z. Peric, "Nonlinear long-term prediction of speech based on truncated Volterra series," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 20, no. 3, pp. 1069-1073, Mar. 2012.

[7] Y. Medan, E. Yair and D. Chazan, "Super Resolution Pitch Determination of Speech Signals," *IEEE Transactions on Signal Processing*, vol. 39, no. 1, pp. 40-48, Jan. 1991.

[8] W. C. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*. *John Wiley & Sons*, New Jersey, USA, 2003.

[9] P. Kroon and B. S. Atal, "On Improving the Performance of Pitch Predictors in Speech Coding Systems," in *Advances in Speech Coding*, B. S. Atal, V. Cuperman and A. Gersho edition, Kluwer Academic Publishers, Norwell, MA., pp. 321–327, 1991.

[10] K. N. Ramamurthy and A. Spanias, *MATLAB Software for the Code Excited Linear Prediction Algorithm: The Federal Standard-1016*, *Synthesis Lectures on Algorithms and Software in Engineering. Morgan & Claypool Publishers*, 2011.