A NEW LINEAR PREDICTIVE METHOD FOR COMPRESSION OF SPEECH SIGNALS

Paavo Alku and Susanna Varho

University of Turku
Lab. of Electronics and Information Technology,
Vesilinnantie 5, FIN-20014 Turku, FINLAND
e-mail: Paavo.Alku@hut.fi

ABSTRACT

A new linear predictive method is presented in this study. The method, Linear Prediction with Linear Extrapolation (LPLE), reformulates the computation of linear prediction by combining the preceding values of sample x(n) into consecutive sample pairs (i.e., x(n-2i), x(n-2i+1)). Each of these pairs determines a regression line the value of which at time instant *n* is used as a data sample in the prediction. The optimal LPLE-predictor is obtained by minimizing the square of the prediction error using the autocorrelation method. The rationale for the new method is the fact that LPLE yields an all-pole filter of order 2*p* when the number of unknowns in the normal equations equals *p*. Therefore the new all-pole modeling method can be used in speech coding applications. Preliminary experiments of the present study show that LPLE is able to model speech spectra more accurately in comparison to conventional linear prediction in the case when a very small number of prediction parameters is required to be used in order to greatly compress the spectral information of speech signals.

1. INTRODUCTION

Linear prediction (LP) is a technique that is widely used in various areas of speech processing, especially in speech coding. In LP-analysis a speech signal is predicted from its past values using an optimal predictor that minimizes the energy of the prediction error, the residual. In the frequency domain this corresponds to modeling the speech spectrum by an all-pole filter [7]. During the past years many modifications of LP-analysis have been presented. It is, for example, possible to modify the selection of data samples in LP-analysis, e.g. [8]., or to change the error criterion that is used in defining the optimal predictor, e.g. [2]. Linear predictive methods have also been developed by replacing the unit delays of the predictor with first order allpass filters [6], [10]. It is also possible to use LP-based methods that take into account various concepts from psychophysics of hearing by using all-pole modeling of the auditory spectrum [4].

Even though many new linear predictive algorithms have been developed during the past years they have not been applied very much in low bit rate speech coding. The focus in speech coding research has been to develop novel algorithms for quantization of the residual using, for example, the multipulse excitation, e.g. [1], or CELP-coding, e.g. [5]. In the present study we propose a new method, named Linear Prediction with Linear Extrapolation (LPLE), which aims at modifying conventional linear prediction especially for speech coding applications. The idea is to reformulate the computation of linear prediction so that an optimal FIR-predictor of order 2*p* could be determined from *p* numerical values.

2. METHOD

Let us denote a sample to be predicted by x(n) as shown in Fig. 1. The idea in the LPLE-method is to predict x(n) by using a known number, denoted by *p*, of sample pairs that occur before time instant *n*. Each sample pair is connected by a line as shown in Fig. 1. The equation for the line that connects two consecutive samples x(n-2i) and x(n-2i+1) can be expressed as follows (time variable is denoted by *k*)

$$f_i(k) = [x(n-2i+1) - x(n-2i)](k - n + 2i) + x(n-2i) \quad (1)$$

Each of the lines that are determined by two consecutive samples are then used to compute the value of the line at time instant *n*, i.e., linear extrapolation is used. By combining all the sample pairs *p* extrapolated values are obtained at time instant *n*. A linear combination of these values are then used to form a prediction for x(n). By denoting the coefficients of the LPLE-predictor by a(i), where $1 \le i \le p$, the following expression is obtained for the predicted value of sample x(n):

$$\hat{x}(n) = \sum_{i=1}^{p} a(i)\left\{2i[x(n-2i+1) - x(n-2i)] + x(n-2i)\right\} \quad (2)$$

Expression for the prediction error can now be presented as follows:

$$e(n) = x(n) + \sum_{i=1}^{p} a(i)\left\{2i[x(n-2i+1) - x(n-2i)] + x(n-2i)\right\} \quad (3)$$

Minimizing the energy of the prediction error with the autocorrelation criterion [7] yields the following normal equations, where the autocorrelation function of x(n) is denoted by R(n):

$$\sum_{i=1}^{p} a(i)\ \{(-4ij + 2i)\, R(2i-2j-1) + (8ij - 2i - 2j + 1)\, R(2i-2j) + (-4ij + 2j)\, R(2i-2j+1)\ \}$$
$$= -2j\, R(2j-1) + (2j-1)\, R(2j), \quad 1 \le j \le p \quad (4)$$

By solving a(i) from Eq. 4 the following transfer function is obtained for the optimal LPLE-predictor:

$$H(z) = 1 + \sum_{i=1}^{p} 2i\, a(i)\, z^{-2i+1} + (1-2i)a(i)z^{-2i} \qquad (5)$$

It is worth noting that Eq. 5 determines an FIR-filter, which is of order 2p. However, the transfer function is obtained from p different values of a(i) as shown by Eq. 4. Hence, in the proposed LPLE-method the order of the predictor is twice the number of unknowns in the normal equations.

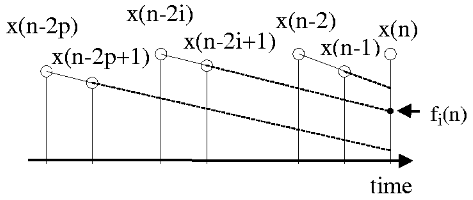

Figure 1: Computation of data samples for prediction of x(n). Each sample pair x(n-2i) and x(n-2i+1) determines a line that is used to compute a linearly extrapolated value, $f_i(n)$, at time instant n.

3. QUANTIZATION OF LPLE-PARAMETERS

In the conventional LP-analysis [7] the order of the optimal FIR-filter, the predictor, equals the number of unknowns in the normal equations. This implies that in speech coding applications based on the conventional LP-analysis with scalar quantization one has to transmit information of p real numbers in order to synthesize the all-pole filter of order p. In the proposed LPLE-method the information that is required in order to synthesize the all-pole filter of order p (i.e, the inverse of the transfer function of Eq. 5) consists of $p/2$ real numbers. Hence, the new method is feasible to be used in speech coding applications where the all-pole filter that models the speech spectrum needs to be quantized with a very small amount of parameters.

Quantization of the filter coefficients of the conventional LP-analysis has been a goal of intensive research, e.g. [11]. Quantization of the LP-parameters can de done using, for example, reflection coefficients, log area ratios (LARs), or line spectrum pairs (see [9] for a review). Quantization of the LPLE-parameters directly (i.e., coefficients a(i), $1 \leq i \leq p$, that are obtained by solving Eq. 4) yields poor matching between the original speech spectrum and its all-pole model if low bit rate for the transmission of the LPLE-predictor is required. Therefore, the following straightforward method was used in this preliminary study for robust quantization of the LPLE-parameters. First, an intermediate FIR-filter is determined using coefficients a(i), $1 \leq i \leq p$, that are obtained after solving Eq. 4. The transfer function of this intermediate FIR-filter is determined as follows:

$$H_i(z) = 1 + \sum_{i=1}^{p} a(i) z^{-i} \qquad (6)$$

Second, the intermediate FIR-filter is quantized using LARs [3]. The purpose of this stage is to compress the original LPLE-information that consists of p real numbers a(i), $1 \leq i \leq p$, to p codewords the lengths of which vary from six to three bits as explained in [3]. Third, the quantized version of the intermediate FIR-filter is obtained by decoding the codewords according to [3]. Fourth, the final quantized version of the LPLE-predictor is obtained using Eq. 5 and the decoded coefficients a(i), $1 \leq i \leq p$, computed in the previous stage.

4. RESULTS

The developed new all-pole modeling technique was compared to the conventional LP-analysis by analyzing voices produced by two female and four male speakers. All the subjects were native speakers of Finnish. The utterances analyzed consisted of four vowels (/a/, /e/, /o/, /ä/), one nasal (/n/) and one fricative (/s/). Each of the signal was analyzed using both the conventional LP-analysis and the proposed LPLE-method. Both of the analyses were computed using the autocorrelation method together with Hamming windowing and the block length of 160 samples (20 ms). First order FIR with its zero at z=0.86 was used as a pre-emphasizer in conventional linear prediction. The number of unknowns in the normal equations (i.e., parameter p in Eq. 4) equaled five. This implies that the order of the conventional LP-filter was equal to five whereas the order of the LPLE-filter equaled ten. However, the information that is required to transmit these two filters consisted in both cases of five real numbers that were quantized with the same bit rate (26 bits per frame). Quantization of both the conventional LP-filter and the intermediate LPLE-filter given in Eq. 6 was done by applying the same procedures used in the quantization of the first five LAR-values of the RPE-LTP-coder of the GSM-system [3].

Examples of the all-pole spectra given by the two predictive methods are shown in Fig. 2, 3, and 4. It can be seen from these graphs that modeling of the speech spectrum in a very compressed form by using only five numerical values can be done more accurately by the proposed LPLE-method than by conventional linear prediction. The all-pole filter of order five given by conventional linear prediction models in general only the over-all structure of the speech spectrum. In the case of LPLE the corresponding all-pole filter of order ten (which is quantized using five parameters) is able to match the formant structure much more accurately. This can be seen in Table 1 that lists the number of formants (i.e., local spectral resonances) found by the conventional LP-analysis and LPLE for the analyzed utterances of each speaker. From this table it can be observed that LPLE is able to find formants clearly more often than conventional linear prediction.

5. CONCLUSIONS

According to our experiments the new predictive method can be applied effectively in applications where all-pole models of speech spectra need to be presented in a very compressed form.

In comparison to the conventional LP-analysis the proposed LPLE-method is able to yield much more accurate models especially for the formant structure of speech in the case when the number of unknowns in the normal equations is small (i.e., p is between 1 and 5). However, when the number of parameters to determine the all-pole filter is larger the differences between the two methods become small. Spectral modeling computed by conventional linear prediction is in general slightly better than modeling yielded by LPLE when the order of the conventional LP-filter is large enough to match the formant structure properly (e.g., $p > 12$). Finally, it is worth noting from Eq. 4 and 5 that the LPLE-method does not always give a minimum phase predictor. This implies that stability of the corresponding all-pole filter can not be guaranteed. However, in the experiments of the present study an unstable LPLE-filter was never obtained.

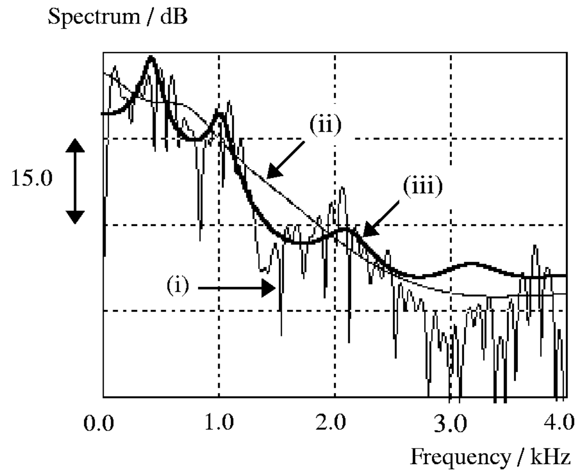

Figure 2: Spectra of the vowel /a/, male speaker, p=5: (i) FFT-spectrum, (ii) quantizated all-pole spectrum given by conventional LP, (iii) quantizated all-pole spectrum given by LPLE.

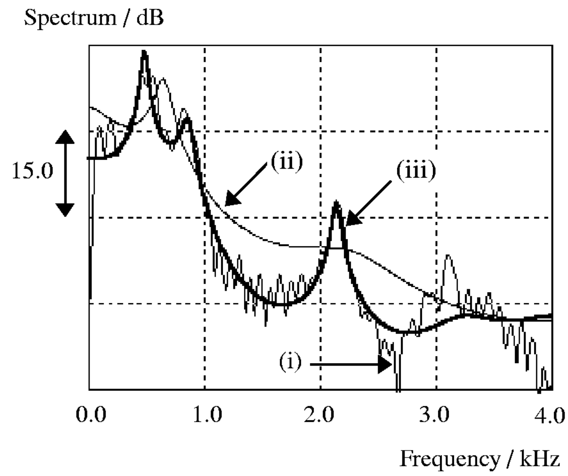

Figure 3: Spectra of the vowel /o/, male speaker, p=5: (i) FFT-spectrum, (ii) quantizated all-pole spectrum given by conventional LP, (iii) quantizated all-pole spectrum given by LPLE.

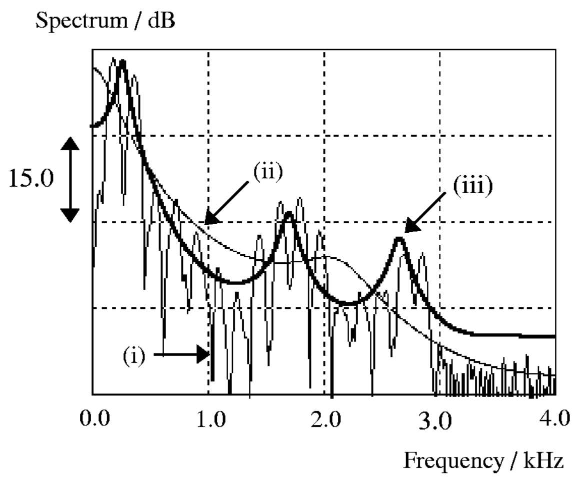

Figure 4: Spectra of the nasal /n/, female speaker, p=5: (i) FFT-spectrum, (ii) quantizated all-pole spectrum given by conventional LP, (iii) quantizated all-pole spectrum given by LPLE.

	/a/		/e/		/o/		/ä/		/n/		/s/	
	LPC	LPLE	LPC	LPLE	LPC	LPLE	LPC	LPLE	LPC	LPLE	LPC	LPLE
F_1	2	4	1	3	2	4	1	3	1	4	2	5
F_2	2	3	2	5	1	3	1	3	1	3	0	0
M_1	1	4	1	3	1	4	1	2	2	3	2	4
M_2	2	4	1	3	2	4	2	3	1	3	2	4
M_3	1	3	1	4	1	3	1	5	1	3	2	4
M_4	2	4	1	2	2	4	1	5	1	2	0	0
Tot	10	22	7	20	9	22	7	21	7	18	8	17

Table 1: Number of local spectral resonances (formants) found from the all-pole spectra given by conventional LPC and LPLE for six utterances (/a/, /e/, /o/, /ä/, /n/, /s/) produced by two female (F_1, F_2) and four male (M_1, M_2, M_3, M_4) speakers.

ACKNOWLEDGMENTS

This study was supported by the Academy of Finland.

6. REFERENCES

1. Atal, B.S. "High-quality speech at low bit rates: Multipulse and stochastically excited linear predictive coders", Proc. ICASSP, pp. 1681-1684, 1986.

2. El-Jaroudi, A., and Makhoul, J. "Discrete all-pole modeling", IEEE Trans. Signal Proc., Vol. 39, No. 2, pp. 411-423, 1991.

3. GSM 06.10, Version 4.0.0, Date: Oct-1992, Released: July 1, 1993, "European digital cellular telecommunication system (phase 2); GSM Full Rate Speech Transcoding", ETSI.

4. Hermansky, H. "Perceptual linear predictive (PLP) analysis of speech", J.Acoust.Soc.Am.,Vol. 87, No. 4, pp. 1738-1752, 1990.

5. Kroon, P., and Atal, B.S. "Quantization procedures for the excitation in CELP coders", Proc. ICASSP, pp. 1649-1652, 1987.

6. Laine, U. "Generalized linear prediction based on analytic signals", Proc. ICASSP, pp. 1701-1704, 1995.

7. Makhoul, J. "Linear prediction: A tutorial review", Proc. IEEE, Vol. 63, No. 4, pp. 561-580, 1975.

8. Miyoshi, Y., Yamata, K., Mizoguchi, R., Yanagida, M., and Kakusho, O. "Analysis of speech signals of short pitch period by a sample selective linear prediction", IEEE Trans. Acoust. Speech Signal Proc., Vol. 35, No. 9, pp. 1223-1240, 1987.

9. Pan, J., and Fischer, T. "Vector quantization of speech line spectrum pair parameters and reflection coefficients," IEEE Trans. Speech Audio Proc., Vol. 6, No. 2, pp. 106-115, 1998.

10. Strube, H.W. "Linear prediction on a warped frequency scale", J. Acoust.Soc.Am., Vol. 68, No. 4, pp. 1071-1076., 1980.

11. Viswanathan, R., and Makhoul, J. "Quantization properties of transmission parameters in linear predictive systems," IEEE Trans. Acoust. Speech Signal Proc., Vol. 23, No. 3, pp. 309-321, 1975.