Mansoura Engineering Journal

Volume 7 | Issue 2 Article 3

7-3-2021

An Improved Sequential Adaptive Method for Linear Prediction of Speech Signals.

Fayez Zaki
Assistant Professor, Electrical Communication and Electronics Engineering Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt., fwzaki@mans.edu.eg

Rasheed El-Awady
Electrical Communication and Electronics Engineering Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt.

Recommended Citation
Zaki, Fayez and El-Awady, Rasheed (2021) "An Improved Sequential Adaptive Method for Linear Prediction of Speech Signals.," *Mansoura Engineering Journal*: Vol. 7 : Iss. 2 , Article 3.
Available at: https://doi.org/10.21608/bfemu.2021.181526

AN IMPROVED SEQUENTIAL ADAPTIVE METHOD FOR LINEAR PREDICTION OF SPEECH SIGNALS

FAYEZ W. ZAKI and RASHEED M. EL-AWADY
Elect. Communication and Electronics Dept., University of El-Mansoura

ABSTRACT:

Much interest has recently been shown in sequential adaptation methods for APC systems. Such methods offer a trade-off between speed of convergence and misadjustment. Inadequacies in the former become most apparent when the speech waveform changes abruptly following certain consonants or unvoiced utterances. This paper offers a method of speeding up convergence at these points by causing the adaptive filter to operate on samples one block ahead of those to be transmitted. Additional time is thus given for adaptation at these critical points.

INTRODUCTION:

A commonly adopted arrangement for residual-excited adaptive predictive coding(APC) system is shown in Fig.1. The speech samples S(n) enter an adaptive filter directly and a "copy" filter via a delay of N samples(typically 100 or 200 samples). The adaptive filter aims to adapt its transfer function to minimise the mean-square value, ε , of its output E(n) over a block of N samples. Its coefficients are then passed over to the copy filter which is used to generate a "residual" signal R(n). The copy filter remains fixed between transfers of coefficients and reprocesses the same block of samples from which its coefficients were calculated by the adaptive filter. R(n) is encoded as $\hat{R}(n)$ (sometimes with some feedback of quantisation error, not shown)for transmission as a low bit-rate representation of the signal S(n). Copy filter coefficients are also transmitted to allow an acceptable version of the original speech to be regenerated by an inverse filter at the receiver.

With residual excited APC systems the encoder for R(n) is generally some form of waveform encoder, e.g, an adaptive delta modulation, operating perhaps on sub-band of the signal. An appropriate set of coefficients should flatten the spectrum of $R(n)$ as well as minimising its mean-square value and these properties are required for efficient operation of the encoder and successful regeneration of speech at the receiver. Hence the coefficients produced by the adaptive filter must be as accurate as possible.

SEQUENTIAL ADAPTATION TECHNIQUES:

Sequential or "sample-by-sample" adaptation techniques differ from the generally more accurate "block" methods in that each sample is processed as it arrives by a very simple algorithm, producing a small modification to the current set of filter coefficients. The calculation of autocorrelation coefficients and matrix manipulation as required by the Autocorrelation and Covariance methods [1,2] is thus eliminated. Typical of such techniques is the LMS algorithm [3] which modifies the vector of coefficients $\underline{a}(n)$, by the following equation:

$$\underline{a}(n+1) = \underline{a}(n) - \mu \hat{\underline{\nabla}}(n) \qquad (1)$$

where μ is a positive constant and $\hat{\nabla}(n)$ is an approximation to the gradient vector:

$$\underline{\nabla}(n) = \left[\frac{\partial \mathcal{E}}{\partial a_1} \quad \frac{\partial \mathcal{E}}{\partial a_2} \quad \ldots\ldots\ldots \quad \frac{\partial \mathcal{E}}{\partial a_m} \right]^T \qquad (2)$$

Equation (1) modifies $\underline{a}(n)$ along the negative gradient of $\mathcal{E}$, thus aiming to minimise $\mathcal{E}$ by "steepest descent". The approximation used for $\hat{\underline{\nabla}}(n)$ is obtained by approximating $\mathcal{E}$ by $E(n)^2$, i.e. the squared value of a single output sample rather than an average over a number of samples. Consequently the method is referred to as a "noisy steepest descent" method.

This algorithm, originally proposed for ladder type filter structures has been modified in various ways [4,5,6,7] to be applicable to lattice and other structures, which are preferable for speech. Zaki and El-Awady [6] have proposed a direct application of equation (1) to a lattice structure where $\underline{a}(n)$ becomes the vector of lattice coefficients $\underline{k}(n)$ and $\hat{\nabla}(n) = \partial(E(n)^2/\partial\underline{k}(n)$ is approximated by

$$\hat{\underline{\nabla}}(n) = 2E(n) \left[b_0(n-1) \quad b_1(n-1) \quad \ldots\ldots\ldots \quad b_{m-1}(n-1) \right]^T \qquad (3)$$

with $b_i(n)$ equal to the ith backward error as defined by Makhoul [8]. Equation (3) defines the "end-point" method which has been found so far to give better results for speech. The technique described in this paper is applicable to all the sequential adaptation systems referred to above.

THE LOOK-AHEAD METHOD:

A difficulty with the use of equation (1) lies with the choice of scaling constant μ. If μ is large, then large changes in coefficients can occur from one sample to the next and adaption can be fairly rapid. However, misadjustment due to the noisy estimate of the gradient will be correspondingly large, leading to wide fluctuations about the steady state solution, a higher than necessary mean-square output and the possibility of instability. On the other hand, a small value of μ will result in a slow rate of convergence especially when the short term frequency spectrum of the speech signal changes rapidly as illustrated in Fig. 3. This waveform was produced by a male speaker saying /POLICEMAN/ and shows the transition from unvoiced to voiced speech at the beginning of /ɔ/ and from nasal to voiced at the beginning of /a/.

Techniques exist for varying the value of μ as adaption proceeds but the fundamental problem still remains. The look-ahead method offers a solution based on a modification to the basic system of Fig.1, presented as Fig. 2. $S(n)$ now passes to the APC transmission system through a further delay D_2, of N samples where the input to the adaptive filter is able to by-pass the delay by means of a switch marked T. With the switch in position A, as shown, the modified system operates in exactly the same way as the original, albeit on a delayed version of $S(n)$. However, when the switch is moved to B, the adaptive filter is effectively receiving data N samples ahead of the signals passing to the transmitter. Its adaption is therefore carried out on this advance data and not on the block of samples that will be transmitted with the resulting coefficients. Switch T must be set to position B only at appropriate times.

In the current application of the look-ahead, the decision as to when to operate T is made on the basis of an estimate of the input signal level. The switch is normally at A and is allowed to change synchronously with the transfer of coefficients to the copy filter. The object is to set the switch to B when the short-term energy of the signal stored in the shift register D_2, and therefore about to pass into the adaptive filter, is below a threshold which approximately distinguishes voiced and unvoiced speech. Clearly, when the signal level is almost zero, for example at some "stop" consonants, there is nothing to be gained by adapting the filter. When the signal level is significant but below the threshold, as may happen with unvoiced and nasal utterances, the coefficients produced by the adaptive filter can be incorrect for the transmission of these samples. However, for such low level signals, correctly adapted coefficients would not in any case have reduced the level substantially, especially as these utterances generally contain less redundancy than the vowels. Their low amplitudes are easily followed by the waveform encoder and since the receiver is residual excited, the fact that their spectra have not been correctly flattened will be compensated in the parameters of the synthesis filter. Sub-band coding techniques which generate artificial excitation bands by spectral distortion will suffer some theoretical loss of quality in these artificial bands which should not be too significant in practice.

The reason for operating switch T in the way described is to allow the adaptive filter to foresee the onset of a voiced signal and to start adapting to it earlier than would otherwise be possible. As soon as the voiced signal is detected by the threshold being exceeded, T reverts to position A and the system proceeds as normal. This means that the block of samples which has just been processed by the adaptive filter passes through it again, this time from the delaying shift register D_2 . The filter adapts twice to this block of samples thus allowing it additional time for convergence from a better starting point.

Note that switch T affects only the adaptive filter input and not the sequence of samples for transmission. The power measurement for the threshold detector may be made on the input signal by an analogue circuit. Only a true/false decision is required every N samples. Alternatively, an approximation may be obtained by summing or digitally low-pass filtering squared samples. This measurement of power is also useful for calculating an appropriate value for u which may then be allowed to vary with signal level.

RESULTS:

The technique described in this paper has been tested in simulation on stored digitised speech one section of which is shown in Fig. 3. The speech bandlimited to 5 KHz and sampled at 10 KHz, was subject to end-point adaptation with and without look-ahead for comparison. The value of μ in each case was allowed to vary with signal level and the threshold for the look-ahead method was set, after some experimentation, at 1.4% of the maximum signal level. The same data was also processed by the Covariance method. In all cases the block size N was 200 samples.

Results for the section /POLICEMAN/ are presented in Fig.4 in the form of a graph showing variations of formant frequency (as indicated by pole positions of the synthesis filter) with sample number. The point of most interest occurs at the onset of /a/ at about sample 2800 (marked with an arrow). Previous to this point, switch T is in position B and the adaption curves for end-point adaptation with and without look-ahead diverge. At sample 2800 it can be seen that the look-ahead method has given a set of pole positions very close to those of the Covariance method whereas the same adaptation method without look-ahead is still adapting towards these pole positions. Taking the Covariance method as a standard it therefore appears that the look-ahead technique has produced a more accurate set of filter coefficients at the start of this voiced utterance at the expence of a less than optimally adjusted filter for the transmission of the low-level nasal /m/.

At the beginning of /POLICEMAN/, the threshold was exceeded not only by /ɔ/ but by the preceding unvoiced signal. The effect is not very significant at this point. The selection of a threshold level is not critical for some improvement to be achieved, but a means of setting the level for optimal effect is still being researched. A refinement being investigated is the replacement of the threshold measurement by a voiced/unvoiced decision made by detecting zero crossing.

CONCLUSION:

A method is presented for improving the performance of sequential adaption techniques when used for residual-excited linear predictive coding in low bit-rate speech transmission systems. A more accurate set of copy filter coefficients are produced for the onset of high-amplitude voiced utterances, at the expense of a less than optimally adjusted filter for preceding low-amplitude unvoiced or nasalised speech. The method is applicable to most sequential adaptation techniques and has been illustrated for one such technique being investigated by the authors.

REFERENCES:

1. Makhoul, J., "Linear Prediction: A Tutorial Review",Proc. IEEE, Vol.63,pp561-580, April 1975.
2. Atal,B. S. and Hanauer,S. L. , "Speech Analysis and Synthesis by Linear Prediction of Speech Wave", J.Acoust.Soc.Am., Vol.50, pp637-655, 1971.
3. Widrow, B. and McCool,J. M., "A Comparison of Adaptive Algorithms Based on the Method of Steepest Descent and Random Search", IEEE Trans. Antenna and Propagation, Vol. AP-24, Sept.1976.
4. Griffiths, L., "A Continuously-Adaptive Filter Implemented as a Lattice Structure", Proc. 1977 IEEE Int.Conf. on Acoust. Speech and Signal Processing Vol. ASSP-77CH197-3,pp 683-686, May 1977.
5. Gibson, C. J. and Haykin, S., "Learning Characteristics of Adaptive Lattice Filtering Algorithms", IEEE Trans.Acoust. Speech and Signal Processing, Vol. ASSP-28, Dec. 1980.

6. Zaki, F. W. and El-Awady, R. M., "An Adaptive Algorithm for Sequential Adaptive Predictive Coding of Speech Signals" Bulletin of the Faculty of Engineering, El-Mansoura University, Vol. 7, No. 1, E 21-E 28, June 1982.

7. Ching, P. C. and Goodyear, C. C., "LMS Algorithm for Sequentially Adapting all-Zero Digital Filters in Modified Cascaded Form", IEE Electronic Letters, Vol. 16, No. 7, PP270-271, March 1980.

8. Makhoul, J. "Stable and Efficient Lattice Methods for Linear Prediction" IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-25, pp 423-428, Oct. 1977.

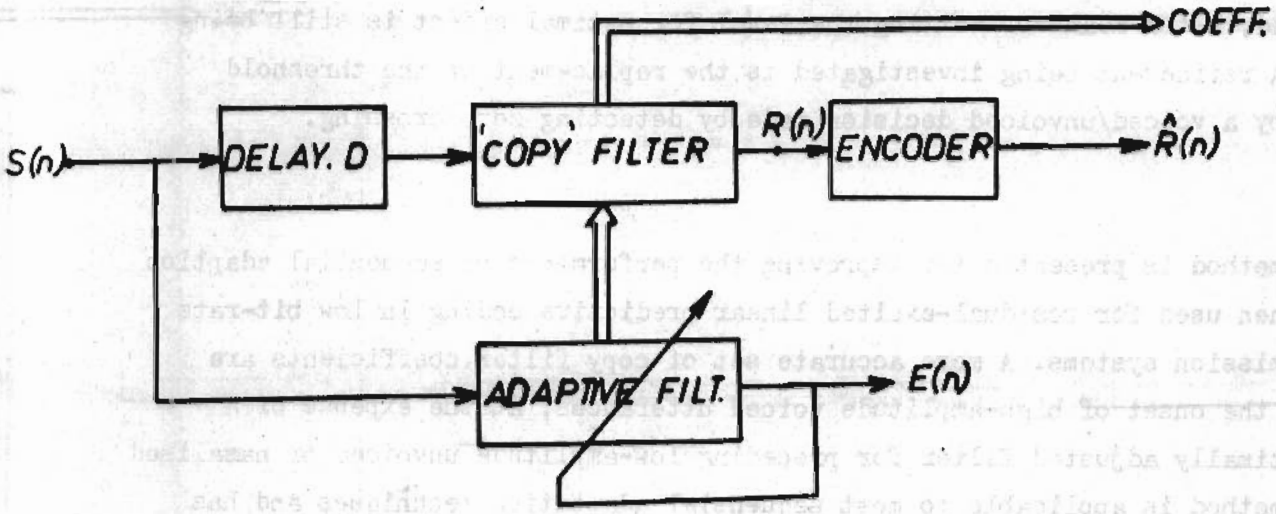

Fig. 1, Residual-Excited Adaptive Predictive Coding System.

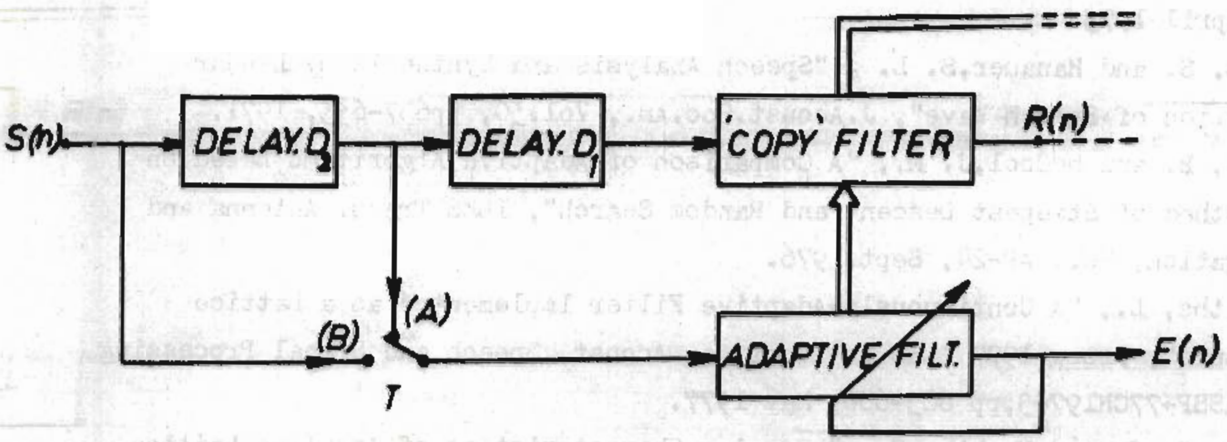

Fig. 2, "Look-Ahead" Arrangement.

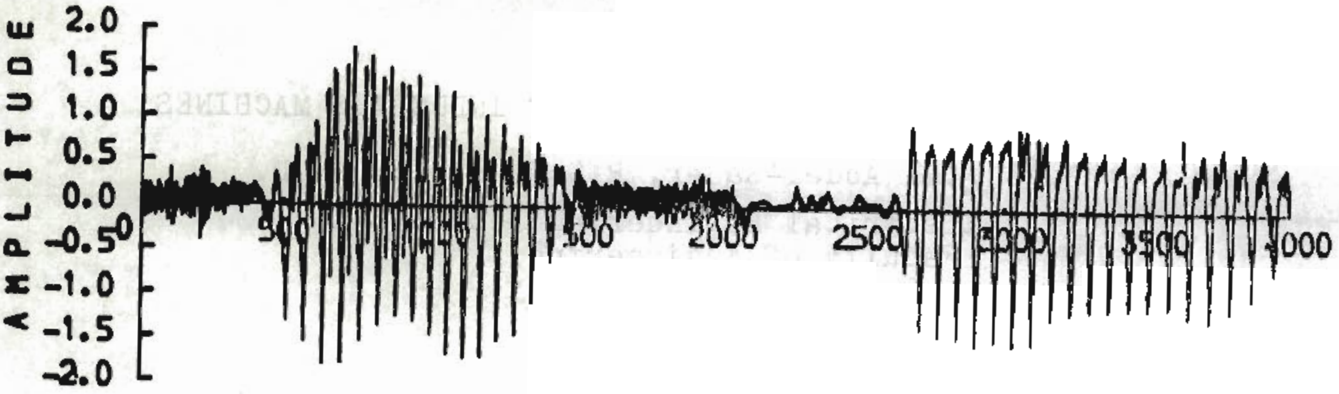

Fig. 3, Acoustic Waveform For /POLICEMAN/

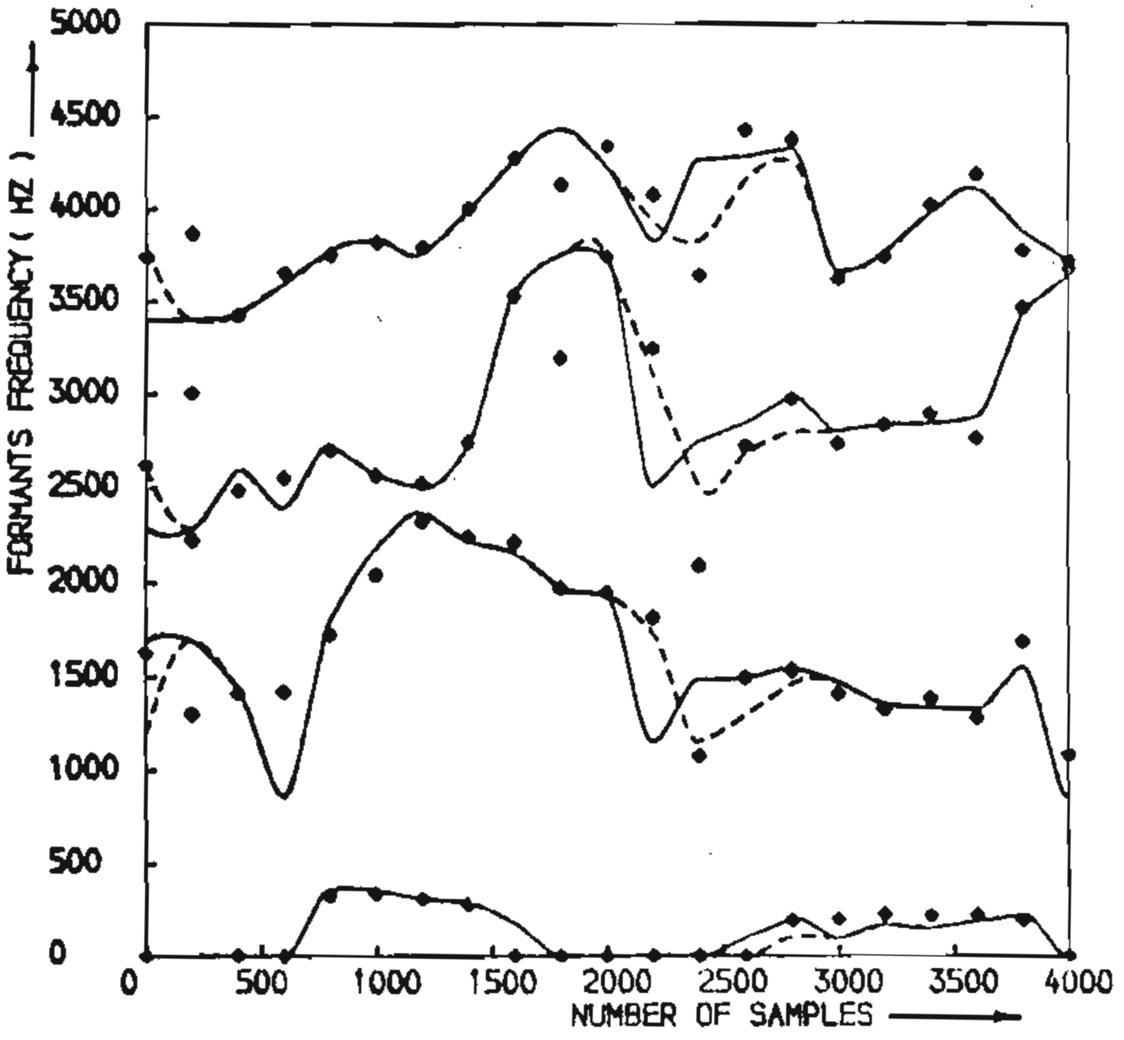

Fig. 4, Formant Tracking For /POLICEMAN/.