SPEECH ENHANCEMENT USING MULTI-PULSE EXCITED LINEAR PREDICTION SYSTEM

K.K. PALIWAL

Computer Systems and Communications Group
Tata Institute of Fundamental research
Homi Bhabha Road, Bombay 400005, India

ABSTRACT

The multi-pulse excited linear prediction (MPELP) system is proposed for speech enhancement. It is shown that for successful enhancement of speech the error-weighting filter should not be used in the MPELP system. A new method (the constrained forward-backward correlation prediction method) is proposed for accurate estimation of LP coefficients from noisy speech. This method guarantees the stability of the estimated all-pole filter which is an important prerequirement for the MPELP system. It is shown that the MPELP system can improve the SNR of 0 dB speech by as much as 5.4 dB.

INTRODUCTION

The problem of enhancing speech corrupted by additive white noise, when only noisy speech is available, is of considerable importance in many applications. Because of this, many speech enhancement algorithms have been developed in the recent past. These algorithms are described in a recent review article by Lim and Oppenheim [1] who have grouped these algorithms in three categories. The first category consists of speech enhancement algorithms based on short-time spectral amplitude estimation and includes the spectrum-subtraction and the Wiener filtering techniques. The algorithms in the second category exploit the periodicity of voiced speech and are based on comb filtering and adaptive noise cancelling techniques. The third category consists of algorithms which are based on the source-system model of speech production and use speech analysis-synthesis system for enhancement.

In the present paper, we propose speech enhancement algorithm using the multipulse excited linear prediction (MPELP) system, recently proposed by Atal and Remde [2] for speech coding. This algorithm falls in the third category and unlike the spectrum subtraction and Wiener filtering algorithms, it does not require any a priori information about the noise statistics.

The MPELP system has been proposed by Atal and Remde [2] for medium-band speech coding. In this system, the residual signal is modelled by multiple pulses for excitation. Atal and Remde have provided an analysis-by-synthesis algorithm for finding the locations and amplitudes of these pulses. An error weighting filter has been used in this algorithm for improving the subjective quality of speech. In the present paper, we propose to investigate the use of the MPELP system for speech enhancement. A block diagram of this system is shown in Fig. 1.

In the context of speech coding, modeling of the residual signal by the multiple pulses causes distortion in the synthesised speech signal. But in the context of speech enhancement, this modeling serves a useful purpose. It can be looked as a noise reduction filter which accepts the noisy residual signal as its input, gets rid of noise between the multi-pulses and outputs a clean excitation signal in the form of multiple pulses for synthesising speech.

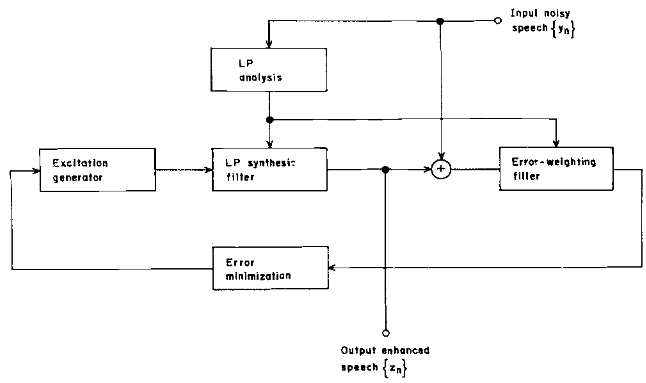

Fig. 1. The MPELP system for speech enhancement.

However, for making the MPELP system work for speech enhancement, we have to solve the following two problems: Firstly, we have to see whether the analysis-by-synthesis algorithm of Atal and Remde is capable enough to compute the locations and amplitudes of the multi-pulses correctly from the noisy speech signal. In case it cannot find multi-pulses correctly for noisy speech, we have to see whether we can modify it to do this job. Secondly, we have to estimate the linear prediction (LP) coefficients correctly from the noisy speech signal. These two problems are dealt with separately in the present paper.

For evaluating the speech enhancement performance of the MPELP system, we have used here two different English sentences: one spoken by a male speaker and the other by a female speaker. Recording of these sentences was done in a quiet room. The speech signal was digitized at a sampling frequency of 8 kHz by a 16-bit analog-to-digital converter.

Noisy speech which is used as the input to the speech enhancement system is obtained by adding an appropriate amount of digitally generated zero-mean white Gaussian noise to the original (clean) speech signal to get the desired input signal-to-noise ratio (SNR). The objective evaluation of the speech enhancement performance is done in terms of SNR of the processed speech. It must be emphasized here that the SNR values do not always give a true picture of the speech enhancement performance. Because of this, informal listening tests were also employed to supplement the objective SNR results.

The paper is organized as follows. In sections 2 and 3, we discuss the problems of estimating multi-pulses and LP coefficients, respectively, from noisy speech. Speech enhancement results are discussed in Section 4 and conclusions in Section 5.

ESTIMATION OF MULTI-PULSES FROM NOISY SPEECH

Here we study the performance of the Atal and Remde's analysis-by-synthesis algorithm in estimating the locations and amplitudes of the multiple pulses from noisy speech. This algorithms requires the estimation of LP coefficients from the noisy speech signal. This is a difficult problem in itself and will be discussed in the next section. In this section, we simplify the problem by using the correct LP parameters computed from the original (clean) speech signal for processing the noisy speech by the MPELP system.

Table I. Speech enhancement performance of the MPELP system using the ideal LP parameters.

Input speech SNR in dB	Output speech SNR in dB	
	With error-shaping filter	Without error-shaping filter
∞	10.75	11.14
10	7.17	10.29
0	-0.66	5.49
-10	-9.96	-3.29

Using these correct LP parameters, we have computed the multi-pulses from noisy speech using the Atal and Remde's algorithm and synthesized speech. The error-weighting filter was used here with a weighting factor r=0.8. Results in terms of SNR values are given in the second column of Table I for different input SNR values. (Here input SNR=∞ means that no noise is added to the original (clean) speech signal.) These objective SNR results indicate that the MPELP system does not improve the quality of speech at all, though the LP coefficients used by it are correct. Informal listening tests also confirm these results.

For this poor performance, we suspected the error-weighting filter used in the analysis-by-synthesis algorithm and, hence, removed it from the system. Noisy speech was processed by the MPELP system once more and the SNR results are shown in the third column of Table I for different input SNR values. It can be seen from this table that with the removal of the error-weighting filter the MPELP system enhances the speech quality of speech significantly. It improves the SNR of 0 dB speech by as much as 5.4 dB. Informal listening by human subjects also confirmed the improved quality of processed speech. It might be noted from Table I that the presence of error-weighting filter in the MPELP system always degrades the SNR performance of the processed speech. This degradation is more for noisy speech having low input SNRs.

Thus, we have seen that the MPELP system can be used for speech enhancement provided the error-weighting filter is removed from the system and the LP coefficients are estimated accurately from noisy speech. We illustrate the results of speech enhancement by the MPELP sysgem for a male speaker in Fig. 2 and for a female speaker in Fig. 3. In these figures, the top part shows the original (clean) speech, the middle part the noisy speech signal at SNR=0 dB and the bottom part the noisy speech signal after enhancement. These figures clearly demonstrate the success of the MPELP system for speech enhancement.

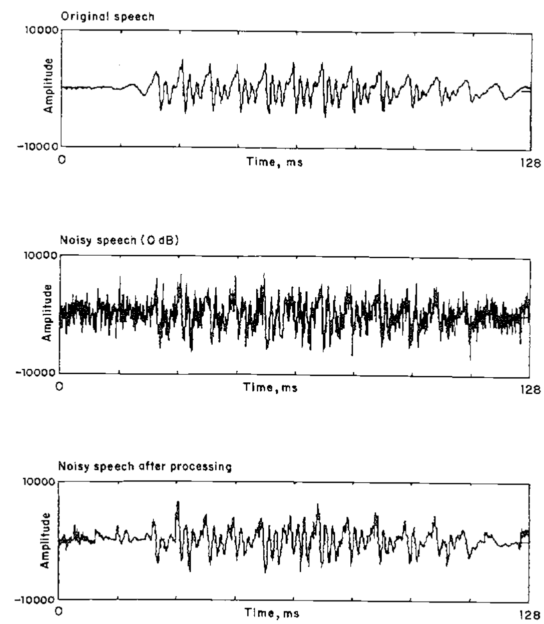

Fig. 2. Illustration of speech enhancement for male speech.

ESTIMATION OF LP PARAMETERS FROM NOISY SPEECH

We have seen in the preceding section that the MPELP system can be employed successfully for speech enhancement provided the LP coefficients are estimated correctly from the noisy speech signal. In this section, we deal with the problem of estimating LP coefficients from noisy speech.

In LP analysis, the speech signal is assumed to be the output of an autoregressive (AR) or an all-pole filter. For noise-free clean speech, the conventional AR spectral estimation methods (such as the autocorrelation method, the covariance method and the Burg method) perform reasonably well [3]. But, when the speech signal is corrupted by the addition of white noise, the AR model is no more valid, and the performance of the conventional methods is poor for the noisy signals

3. 6. 2

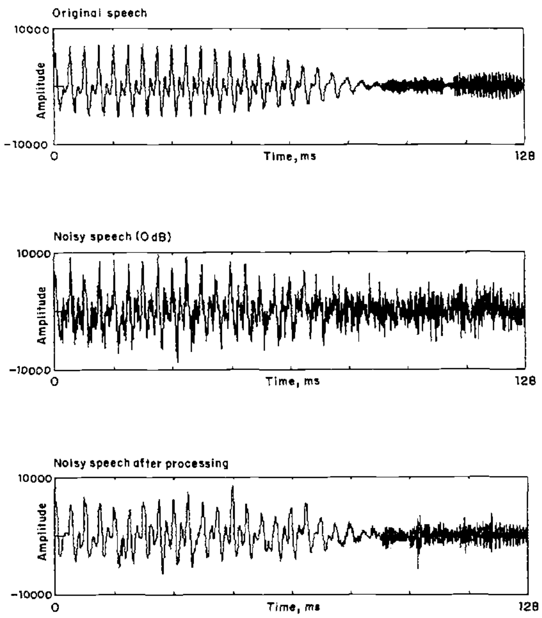

Fig. 3. Illustration of speech enhancement for female speech.

[4,5].

Some methods have been reported in the literature for AR spectral estimation of noisy signals. These methods use p high-order Yule-Walker equations for estimating the parameters of the pth order AR model [6,7]. Recently, Cadzow [8] has proposed the use of an overdetermined set of q(>p) high-order Yule-Walker equations for estimating p AR parameters (or LP coefficients). In Cadzow's formulation, each high-order autocorrelation coefficient is predicted in terms of p preceding autocorrelation coefficients and the AR parameters are obtained by minimizing the total-squared correlation prediction error. Since this method uses the forward prediction of autocorrelation coefficients, we shall refer it as the forward correlation prediction (FCP) method. Codzow [8] has also proposed another method (the forward-backward correlation prediction (FBCP) method) which uses forward as well as backward prediction of autocorrelation coefficients for estimating the AR parameters.

A major problem with the FCP and FBCP methods is that these methods do not ensure the stability of the estimated AR system. For the MPELP system used in the present paper for speech enhancement, this is an important prerequirement.

In the present paper, we propose a new method, the constrained forward-backward correlation prediction (CFBCP) method, for estimating AR parameters from noisy speech. An important feature of this method is that it guarantees the stability of the estimated AR system. Like the FBCP method, this method uses the forward as well as backward prediction of autocorrelation coefficients. But, to ensure stability, it minimizes the total-squared forward-backward correlation prediction error under the constraint of Levinson's recursion relations [3].

In order to illustrate the performance of the CFBCP method, we take here 200 samples of noisy speech at SNR=10 dB of a synthetic vowel /i/ generated at 8 kHz sampling rate and employ p=10 and q=12. Figure 4 shows the true power spectrum of the signal and power spectrum estimates using the conventional autocorrelation method [3] and the CFBCP method. Superior performance of the CFBCP method can be clearly seen from this figure. Detailed results about the CFBCP method are presented in an another paper [9].

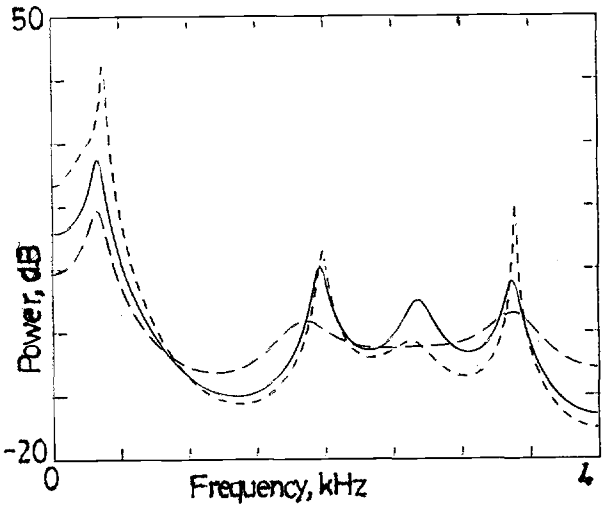

Fig. 4. True spectrum (solid line) of vowel /i/ and its estimates from the autocorrelation method (long-dashed line) and the CFBCP method (short-dashed line). SNR=20 dB.

SPEECH ENHANCEMENT RESULTS FROM THE MPELP SYSTEM

In this section we use the MPELP system shown in Fig. 1 without the error-weighting filter for speech enhancement and employ the CFBCP method described in the preceding section for computing the LP coefficients from noisy speech. The SNR values of processed speech are shown in Table II for different input SNR values. It can be seen from this table that the MPELP system improves the speech quality significantly. For 0 dB input

Table II. Speech enhancement performance of the MPELP system using the CFBCP estimates of LP parameters from noisy speech.

Input speech SNR in dB	Output speech SNR in dB
∞	8.20
10	8.32
0	3.43
-10	-6.59

3. 6. 3

speech, it gives an improvement of about 3.4 dB. The informal listening tests also confirm these results.

It might be noted from Tables I and II that for input SNR=0 dB the MPELP system can improve the speech quality by 5.4 dB by using the ideal LP coefficients (i.e., the LP coefficients derived from the original (clean) speech), while improvement in SNR is only 3.4 dB using the LP coefficients estimated by the CFBCP method from noisy speech. Thus, though the CFBCP method proposed in the present paper does a good job of spectral estimation from noisy speech (as can be seen from Fig. 4) and can be used successfully in the MPELP system to enhance speech, a better method of spectral estimation can improve further the quality of processed speech. There is a scope for improvement in SNR by 2 dB.

CONCLUSIONS

In the present paper, the MPELP system is proposed for speech enhancement. This system can be used successfully to enhance speech provided the error-weighting filter is not used. A new CFBCP method is proposed for accurate estimation of LP coefficients from noisy speech. This method has an important feature that it guarantees the stability of the estimated AR system. It is shown that the MPELP system is capable of enhancing speech by as much as 5.4 dB for highly noisy input speech (with SNR equal to 0 dB or less). Informal listening tests also confirm these objective SNR results.

ACKNOWLEDGEMENT

The author wishes to thank Dr. T. Svendsen of NTH, Trondheim, Norway, for his contribution during the first part of this study.

REFERENCES

[1] J.S. Lim and A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech", Proc. IEEE, Vol. 67, Dec. 1979, pp. 1586-1604.

[2] B.S. Atal and J.T. Remde, "A new model of LPC excitation for producing natural sounding speech at low bit rates", Proc. IEEE Intern. Conf. Acoust., Speech and Signal Process., Paris, 1982, pp. 614-617.

[3] J. Makhoul, "Linear prediction: A tutorial review", Proc. IEEE, Vol. 63, Apr. 1975, pp. 561-580.

[4] M.R. Sambur and N.S. Jayant, "LPC analysis/synthesis from speech inputs containing quantizing noise or additive white noise", IEEE Trans. Acoust., Speech and Signal Process., Vol. ASSP-24, Dec. 1976, pp. 488-494.

[5] S.M. Kay, "The effects of noise on the autoregressive spectral estimator", IEEE Trans. Acoust., Speech and Signal Process., Vol. ASSP-27, Oct. 1979, pp. 478-485.

[6] W. Gersch, "Estimation of the autoregressive parameters of a mixed autoregressive moving-average time series", IEEE Trans. Automatic Control, Vol. AC-15, Oct. 1970, pp. 583-588.

[7] Y.T. Chan and R.P. Langford, "Spectral estimation via the high-order Yule-Walker equations", IEEE Trans. Acoust., Speech and Signal Process., V01. ASSP-30, Oct. 1982, pp. 689-698.

[8] J.A. Cadzow, "Spectral estimation: An overdetermind rational model approach", Proc. IEEE, Vol. 70, Sept. 1982, pp. 907-939.

[9] K.K. Paliwal, "A constrained forward-backward correlation prediction method for AR spectral estimation of noisy signals", to be published.