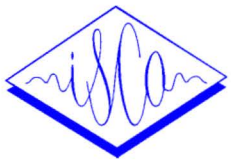

First European Conference on
Speech Communication and Technology
EUROSPEECH '89
Paris, France, September 27-29, 1989

COMBINED OPTIMIZATION OF EXCITATION AND FILTER PARAMETERS IN ANALYSIS-BY-SYNTHESIS CODERS

K.Y.Lee B.G.Evans

Department of Electronic and Electrical Engineering
University of Surrey, Guildford, Surrey, U.K.

ABSTRACT

Analysis-by-Synthesis LPC (ABS-LPC) based speech coding schemes have been very successful at producing good quality speech below 8Kbit/s. In this paper we report on work to further enhance on the capabilities of ABS-LPC schemes, particularly CELP. Specifically, our attention has been focused on re-optimising the time-varying filter of ABS-LPC schemes by introducing a sequential iterative estimation procedure which extracts the best excitation and filter combination. Also, by adopting a pole-zero filter in place of the all-pole LPC filter a more general approach is introduced which gives better coverage of the speech variations. This new coder, ARMA-CELP, was found to give good objective and subjective results when compared with a standard CELP coder.

1. INTRODUCTION

In recent years, efforts to bring the transmission rate of digitally coded speech below 9.6Kbit/s have mainly concentrated on analysis-by-synthesis (ABS) techniques and the extensive use of vector quantization. Most of the ABS schemes employ a time varying linear recursive filter in their synthesis model, and systems that have emerged include MPLPC [1], RPE-LPC [2], CELP[3], SELP [4]. These ABS-LPC systems, while based on the source filter model of speech production, use a limited closed-loop optimization procedure to determine a low bit rate excitation signal which when used to excite the model filter, minimises a perceptually weighted error formed between the original and the synthesized speech samples. It is this limited closed-loop approach which has enabled these coders, especially CELP, to be more successful at bit rates below 7Kbit/s than conventional analysis-and-synthesis systems.

In all the ABS-LPC schemes mentioned above, the time-varying filter parameters are determined only once before the closed-loop excitation sequence optimization. Once the optimum excitation sequence is determined no attempts are made to re-optimise the filter parameters, i.e. the filter optimization procedure is not included in the closed-loop system. In an effort to optimise both the excitation and filter parameters, we have investigated the use of a pole-zero (ARMA) filter inside a CELP coding scheme, and this shall be reported in this paper.

2. SYSTEM DESCRIPTION

A block diagram of a CELP [3] coder is shown in Fig.1. In standard CELP, the basic order of analysis is as follows:

(1) Calculate the short term predictor (STP), H(z), i.e. LPC parameters and form the first residual.

(2) Using the closed-loop or open-loop analysis, calculate the long term predictor (LTP), P(z), or pitch parameters, i.e. delay and gain.

(3) Remove contributions of the LTP and STP memory from the original to form a reference signal.

(4) Estimate the best codebook excitation via a trial and error process to give a minimial error between the reference and the synthetic signal.

(5) Synthesize speech signal by putting back all initial conditions to the LTP and STP with the optimal codebook sequence.

From the outline of the CELP algorithm above, two points can be observed.

(1) The algorithm is not truly analysis-by-synthesis as the STP is fixed for the frame of analysis.

(2) The excitation sequence is only optimal for the particular STP estimated. If the STP is incorrect, then the excitation sequence will be sub-optimal.

The above mentioned deficiencies in the standard CELP algorithm has prompted us to re-examine the CELP algorithm. Ideally, for our CELP coder to be truly ABS, it would be desirable to be able to estimate all the coder parameters, i.e. STP, LTP, excitation, at once. This is obviously very costly in computational terms since standard CELP, which is not full ABS, already requires enormous number of computations (500 MIPS). Therefore, as a compromise, we propose a sequential method where the STP and the STP excitation are iteratively estimated. Although the STP (LPC) filter in normal CELP is usually quite accurate it can fail to be optimal for certain types of input speech signals. For example, (a) when the speech is predominantly nasal in character the LPC analysis will be unable to spectrally match the zeros present in the speech, i.e. the "valleys" in the STP envelope, (b) when significant amount of noise is present in the speech so that the spectral information is degraded. Consequently, we propose the replacement of the LPC STP filter with a pole-zero (ARMA) type filter. This will provide a more general analysis model to give better coverage of the wide variety of speech characteristics, and should perform better under noisy conditions. This new coding system shall be termed ARMA-CELP.

2.1 BLOCK COMPONENT PROCEDURE

As indicated earlier, our goal is to iteratively estimate the STP and the STP excitation sequence until a lower bound is reached. Our new procedure of speech coding follows along the lines as outlined above for CELP, and is briefly as follows:

(1) Extract the excitation sequence from a LPC residual.
(2) Estimate the best ARMA filter with the extracted excitation sequence.
(3) Extract the excitation sequence using the best ARMA filter produced.

Steps (2) and (3) are repeated until convergence is reached, or after a set number of iterations. This iterative block component procedure is depicted in Fig.2. As usual in speech coding, this block component procedure stresses that the error minimisation is accomplished within each block of analysis.

2.2 EXCITATION SEQUENCE

The STP excitation sequence can be modelled in a number of ways. A LTP is usually employed as it very effectively provides the bulk of the excitation during voiced periods of speech. In this study a modified open-loop LTP analysis method [5] is used. This method provides comparable performance to full closed-loop LTP analysis, but without the extra computations involved. The memory of the LTP is provided by the primary excitation which also provides the bulk excitation when the speech is less periodic. This primary excitation can be formed using a variety of methods, e.g. multi-pulse, codebook, etc.. For this study codebook excitation was chosen as it offers good efficiency in transmission terms. A centre clipped gaussian codebook [6] was selected as this is generally reported to give good performance. The optimal codebook entry is given by the sequence which maximises Eqn.1.

$$E_j = \frac{[\sum_{k=0}^{P-1} e_2(k)\, y_j(k)]^2}{\sum_{k=0}^{P-1} y_j(k)} \qquad (1)$$

where $e_2(k)$ is the memoryless reference signal (see Fig.1), and $y_j(k)$ is the j^{th} filtered synthetic reference signal produced by the j^{th} codebook codeword.

2.3 POLE-ZERO FILTER ESTIMATION

In this sections the outline of proposed LPC filter replacement will be described. Let us assume that the speech signal $s(k)$ is described as

$$s(k) = h_A(k) * u(k) \qquad (2)$$

where "*" denotes convolution, and $u(k)$ is the excitation sequence obtained in section 2.1, and $h_A(k)$ is the system impulse response to be estimated. If we assume a general system for speech, then $h_A(k)$ can be characterised by a pole-zero model with transfer function

$$H_A(z) = \frac{G + \sum_{j=1}^{M} \beta_j z^{-j}}{1 + \sum_{i=1}^{N} \alpha_i z^{-i}} \qquad (3)$$

The corresponding difference equation can be written as

$$s(k) + \sum_{i=1}^{N} \alpha_i s(k-i) = Gu(k) + \sum_{j=1}^{M} \beta_j u(k-j) \qquad (4)$$

If $\beta_j = 0$, $j = 1,...,M$, the above becomes our general LPC analysis model. A wide range of parameter estimation techniques can be used to identify the unknown parameters (α_i, β_j) when the excitation $u(k)$ is available. In this study, we have used a Kalman type estimation technique [7].

Let $\mathbf{A}$ be a $(N+M)$ column vector.

$$\mathbf{A} = [\alpha_1, \cdots, \alpha_N, \beta_1, \cdots, \beta_M]^T \qquad (5)$$

Since $\mathbf{A}$ is a constant vector, it can be modelled by the equation

$$\mathbf{A}(k+1) = \mathbf{A}(k) \qquad (6)$$

If the input $u(k)$ were available, which we have from the prior estimation stages, we can form the row vector $\mathbf{C}(k)$,

$$\mathbf{C}(k-1) = [-s(k-1), -s(k-2), \cdots, -s(k-N), u(k-1), \cdots, u(k-M)] \qquad (7)$$

and write the observation equation as

$$z(k) = s(k) = \mathbf{C}(k-1)\mathbf{A}(k) + Gu(k) \qquad (8)$$

Following standard Kalman notation, the identification algorithm then becomes

$$\hat{\mathbf{A}}(k+1) = \hat{\mathbf{A}}(k) + \mathbf{K}(k+1)[s(k+1) - \mathbf{C}(k)\hat{\mathbf{A}}(k)] \qquad (9)$$

$$\mathbf{K}(k+1) = \mathbf{V}_{\tilde{\mathbf{A}}}(k)\mathbf{C}^T(k)[G^2 + \mathbf{C}(k)\mathbf{V}_{\tilde{\mathbf{A}}}(k)\mathbf{C}^T(k)]^{-1} \qquad (10)$$

$$\mathbf{V}_{\tilde{A}}(k+1) = [\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k)]\mathbf{V}_{\tilde{\mathbf{A}}}(k) \qquad (11)$$

where $\mathbf{K}(k)$ is the gain vector and $\mathbf{V}_{\tilde{A}}(k)$ is the error variance matrix.

As evident the technique is therefore sequential in nature, and the model orders, N and M, must be fixed a priori. Thus the procedure starts with an initial set of $\hat{\mathbf{A}}$ and is updated every sample with a gradient heading towards the real A parameters. Usually, all-pole parameters from the first LPC analysis are good candidates as initial estimates of $\hat{\mathbf{A}}(0)$. The estimation is terminated after the last sample. The choice of the initial error variance, $\mathbf{V}_{\tilde{\mathbf{A}}}(0)$ is also critical, and should reflect the degree of confidence placed in the initial choice of $\hat{\mathbf{A}}$. If this initial choice is thought to be good, $\mathbf{V}_{\tilde{\mathbf{A}}}(0)$ should be small so that the estimates change very little. If the initial choice is thought to be poor (or uncertain), $\mathbf{V}_{\tilde{\mathbf{A}}}(0)$ should be chosen large. The initial choices significantly affect the convergence of the algorithm, i.e. the rate at which the gain vector $\mathbf{K}(k)$ tends to zero. Although it would appear to be desirable to choose initial values so that convergence is accelerated, such a procedure could cause the algorithm to converge to a value which is not the correct value of the system. This is related to the phenomenon of divergence in the Kalman filter.

As shown in Fig.2, there is an iteration loop for the excitation estimation and ARMA filter update. Initially, a minimum phase response is used for excitation optimisation. Once an updated ARMA filter is obtained it is then used to improve the

excitation sequence. Usually, this process converge after two or three iterations.

3. SIMULATION RESULTS

This section presents the experimental results obtained for the proposed ARMA-CELP coder. The system configuration parameters are shown in Table 1. For the simulations the number of filter re-optimisation iterations was fixed at 3. As can be expected, the amount of computational time was substantially higher than that of a normal CELP coder. In order to reduce the amount of search time, simplifications were experimented. During periodic regions, the LTP provides most of the excitation, and this is indicated by a large LTP gain term. Thus, if a large LTP gain term was encountered, the codebook search was omitted. Thus, only during the final iteration is the codebook search carried out. As the codebook search make up the bulk of a CELP system this simplification procedure reduced the total amount of computations considerably. The performance as a result of the simplification was not significantly affected.

The average gains for the test material used for the new coder and a similarly configured standard CELP are shown in Table 2. As can be observed, the amount of performance improvement over CELP is only respectable considering the extra amount of computations required. However, it was noticed that the fluctuations in the SNR values for the ARMA-CELP is much more than for CELP, and in some instances the difference between ARMA-CELP and CELP was over 4dB. This indicates that our parameter estimation technique remains to be better "tuned" so that higher SNR's can be better sustained throughout a test sentence. This can be achieved by better choices of initial starting conditions for the estimation algorithm. Another factor highlighting the non-ideal starting conditions is the incidences of unstable pole-zero filters observed in the pole-zero STP. Although the zero filter coefficients, β, can be non-minimum phase, the pole filter coefficients, α, should preferably be minimum phase. The stability of the α coefficients were determined by converting them into the equivalent reflection coefficients (PARCOR), k_i, i.e. stable if $|k_i| < 1.0$. The percentage of unstable pole-zero filters observed was in the order of 10-15% depending on the particular test sentence. When the pole-zero filter was unstable, the coding system reverts back to using the normal LPC coefficients. This relatively high percentage of unstable filters could explain the large variance of the SNR measurements.

In informal subjective listening tests, the processed speech was found to be "fuller" than the CELP equivalent. Also it was noticed that the voiced regions were better reproduced with distinctively sharper processed speech. However, there was still some noticeable background roughness, as was observed in the CELP. Also, during some regions as discussed above, there were some slightly audible spectral distortions. This was probably caused by the discontinuity encountered when the estimation was unable to converge to a stable pole-zero filter.

As an enhancement tool, a pole-zero type post-filter as described in [8] was applied to the processed speech. The attachment of the post-filter produced a significant improvement in the processed speech. The background roughness was substantially reduced, and the slight spectral distortion mentioned earlier was also largely suppressed.

4. CONCLUSION

In this paper we have reported on attempts to further enhance on the capabilities of ABS-LPC schemes, particularly CELP. Specifically, our attention have been focused on re-optimising the time-varying filter by introducing a more general approach by adopting a pole-zero filter instead of an all-pole filter. This new coder, ARMA-CELP, was found to give better objective and subjective performance than standard CELP. The processed speech, especially after post-filtering, was very good quality. However, as was discussed earlier, fine points of the new ARMA-CELP coder still remains to be fully investigated. These, and methods of reducing the complexity, will be the subject of our future work.

References

[1] B.S.Atal, J.R.Remde, "A new model of LPC excitation for producing natural sounding speech at low bit rates", Proc. of ICASSP-82, pp614-617.

[2] P.Kroon, E.F.Deprettere, R.S.Sluyter, "Regular pulse excitation: A novel approach to effective and efficient multi-pulse coding of speech", IEEE Trans. ASSP-34, No.5, pp 1054-1063, 1986.

[3] M.R.Schroeder, B.S.Atal, "Code-Excited Linear Predication (CELP): High quality speech at very low bit rates", Proc.of ICASSP-87, pp 1649-1652.

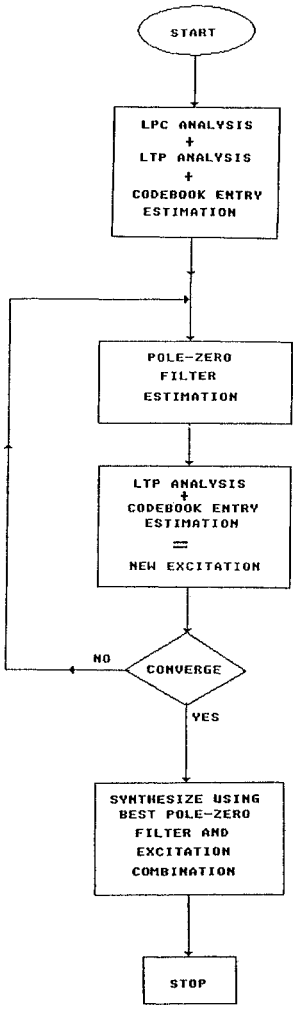

Fig.2: Flow diagram of block component procedure for estimating the best filter and excitation combination.

[4] R.C.Rose, T.P.Barnwell, "The Self-Excited Vocoder: an alternative approach to toll quality at 4800bps", Proc. of ICASSP-86, pp 453-456.

[5] A.Kondoz, "Digital Encoding of Speech Signals at 16 to 4.8Kbit/s", PhD Thesis, Surrey University, U.K., 1988.

[6] P.Kroon, B.Atal, "Strategies for Improving the Performance of CELP Coders at Low Bit Rates", Proc. of ICASSp-88, pp 151-154.

[7] M.Srinath, P.Rajasekaran, "An Introduction to Statistical Signal Processing with Applications", John Wiley and Sons, U.S.A, 1979.

[8] J.Chen, A.Gersho, "A Real-Time vector APC Speech Coding at 4800bps with Apaptive Processing", Proc. of ICASSP-87, pp 2185-2188.

Sampling	8KHz
Frame length	25ms
Subframe length	6.25ms
LTP/update	1-tap/6.25ms
Codebook/update	10bit/6.25ms
LPC order	10-pole
Pole-zero orders	10-pole/4-zero
Weighting factor (w)	0.90

Table 1: Analysis parameters for the ARMA-CELP coder

Coder	Female SEGSNR (dB)	Male SEGSNR (dB)
ARMA-CELP	10.45	10.52
CELP	9.56	9.73

Table 2: SEGSNR values for ARMA-CELP and a CELP equivalent using the analysis parameters of Table 1.

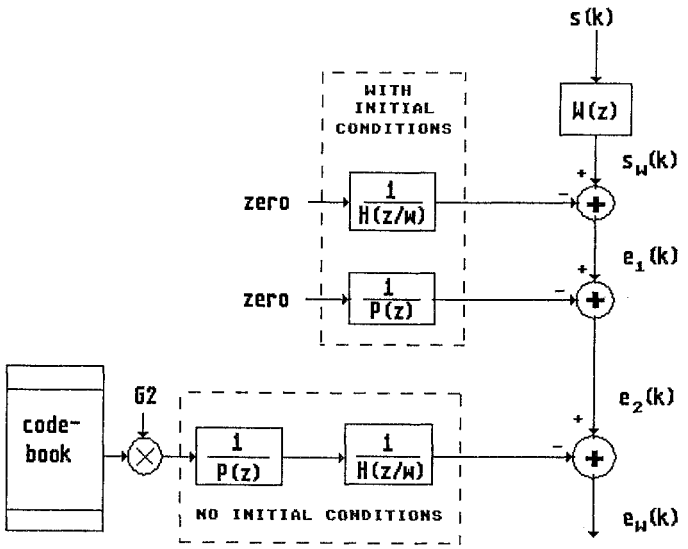

Fig.1a: Block diagram of CELP encoding procedure.

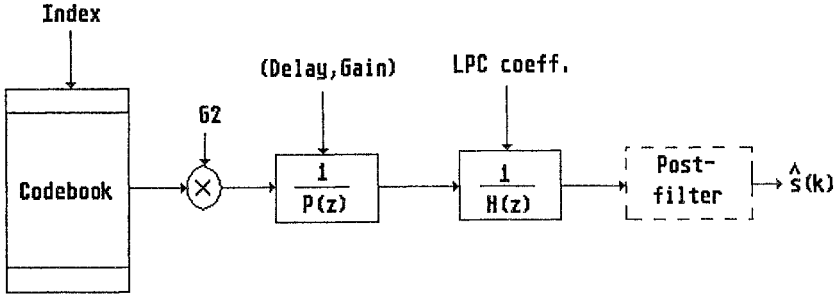

Fig.1b: Block diagram of CELP synthesizer.