N88-25751

REAL-TIME SPEECH ENCODING BASED ON CODE-EXCITED LINEAR PREDICTION (CELP)

WILFRID P. LEBLANC, Carleton University, Canada;
S.A.MAHMOUD, Carleton University, Canada.

CARLETON UNIVERSITY
Department of Systems and Computer Engineering
Ottawa, Ontario
Canada
K1S 5B6
Tel: 1-613-564-6655, 1-613-564-2793

ABSTRACT

This paper reports on the work proceeding with regards to the development of a real-time voice codec for the terrestrial and satellite mobile radio environments. The codec is based on a complexity reduced version of CELP. The codebook search complexity has been reduced to only 0.5 MFLOPS (million floating point operations per second) while maintaining excellent speech quality. Novel methods to quantize the residual and the long and short term model filters are presented.

INTRODUCTION

Since the introduction of CELP [Schroeder 1985], there has been considerable interest in techniques to reduce its complexity, and efficient techniques to encode the long and short term model filters. In Schroeder's and Atal's classic paper, the residual was quantized at only 2 kbps, with no quantization of the remaining parameters (gain, long and short term predictor). It was believed that a high quality voice codec, based on CELP, could be realized at an aggregate rate of 4.8 kbps.

However, the computational complexity of CELP is unwieldy, which hinders the real-time implementation of such a technique. Various authors [Davidson 1986, Trancoso 1986, Chen 1987] have addressed the complexity issue and have arrived at marginally acceptable systems.

In this paper, we propose a new structure— referred to as KELP— which has a greatly reduced complexity. We also discuss novel techniques to encode the long and short term predictor parameters.

The original CELP algorithm is shown in Fig. 1, The filter $A(z)$ is referred to as the short term inverse filter and is given by [Markel 1976]

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}$$

where p is the predictor order and the a_k's are the predictor parameters. The long term inverse filter $B(z)$, is given by

$$B(z) = 1 - \sum_{k=-q}^{q} b_k z^{-(M+k)}$$

where the b_k are the long term predictor parameters and M is the pitch period. The order is large but many of the coefficients are zero. The noise weighting filter $W(z)$ is just

$$W(z) = \frac{A(z)}{A(z/\gamma)}$$

where $\gamma \approx 0.73$.

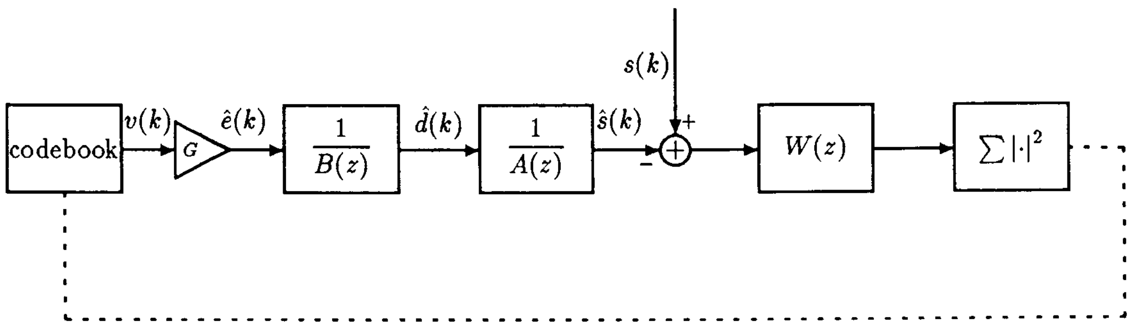

Fig. 1. The Original CELP Search Algorithm

The search complexity for the optimal excitation from the L level, K dimensional codebook is just

$$\begin{aligned} C &= LK(3p + 2q + 2)f_s/K \quad \text{FLOPS} \\ &\approx 300 \quad \text{MFLOPS} \end{aligned}$$

for $L = 2^{10}$, $K = 40$, $p = 10$, $q = 1$, and a sampling frequency (f_s) of 8 kHz.

REDUCED COMPLEXITY CELP (KELP)

In this section we introduce a technique to greatly reduce the computational complexity of the CELP algorithm while maintaining excellent speech quality.

As the first step, we move the location of $W(z)$ and separate the zero state and zero input responses of the long and weighted short term model filter ($A(z/\gamma)$). If the pitch period is greater than the vector dimension ($M > K + q$), the zero state response of $B(z)$ is unity. Thus, we may redraw the structure as shown in Fig. 2. Note that the ordering of G and $A(z/\gamma)$ is not important since we are dealing with the zero state response only.

We note that the structures of Fig. 1 and Fig. 2 are not amenable to fast search non-exhaustive algorithms. By moving the location of the codebook, and quantizing $x(k)$, a non-exhaustive (tree search) is possible. In addition, if the denominator of $W(z)$ is fixed, (as in [Kroon 1986]), no degradation would result. With a variable denominator filter, degradation will result. However, $A(z/\gamma)$ is relatively flat and contains no pitch information. Thus, quantizing $x(k)$ should be very efficient. Forcing the codebook elements to the unit circle implies that minimizing the mean squared error is equivalent to maximizing the correlation

$$G = \sum_{k=0}^{K-1} x(k)\hat{x}(k)$$

over the whole codebook. The modified search procedure is shown in Fig. 3. With the search structured in this manner, we may use a non-exhaustive tree search for the optimum excitation.

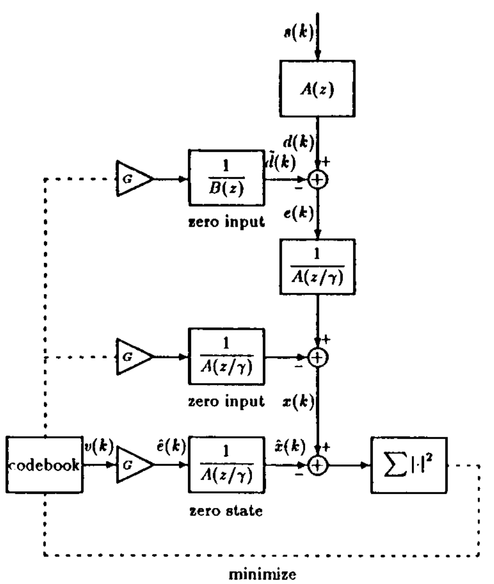

Fig. 2. CELP, Modified Search

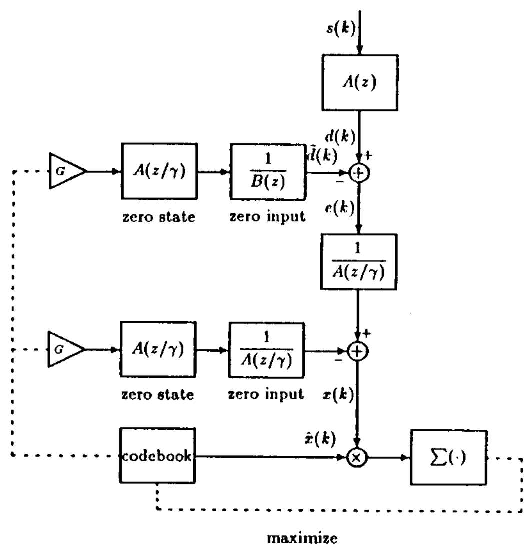

Fig. 3. KELP

The search complexity for this structure is greatly reduced. For a full search code the complexity (for the above parameters) is 8 MFLOPS. By using a binary tree search, or a 32–32 tree search [Makhoul 1985], we obtain complexities of 160 KFLOPS, and 500 KFLOPS respectively. A 32–32 tree search is a good tradeoff between computational complexity and memory requirements.

Typically the pitch predictor is optimized to minimize the open loop residual energy. However, the input to the pitch predictor contains an appreciable amount of quantization noise. By redefining the noise weighting filter we can circumvent this problem. We redefine $W(z)$ such that

$$\tilde{W}(z) = \frac{A(z)B(z)}{A(z/\gamma)}$$

The search procedure with the new noise weighting filter is shown in Fig. 4. Note that the new noise weighting filter emphasizes the pitch information more so than the original weighting filter. Perceptual quality was judged to be almost equivalent.

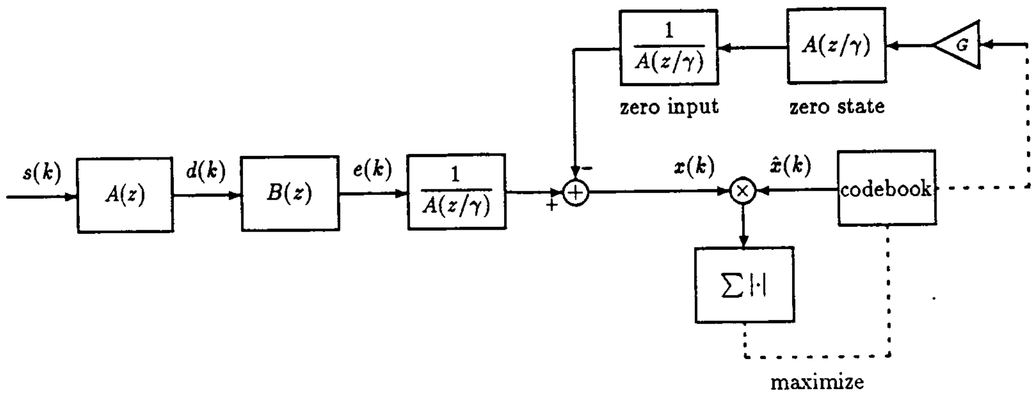

Fig. 4. KELP, Redefined Noise Weighting Filter

The memory complexity ($\mathcal{M}$) and computational complexity ($\mathcal{C}$) for the various search procedures is shown in Table 1. The codebooks were optimized using a closed loop Vector Quantizer (K-means) design algorithm [LeBlanc 1988, Makhoul 1985]. The bit rate for the residual with the above parameters is 2 kbps.

Table 1. Complexity of CELP and KELP

Codec	$\mathcal{M}$ (KWords)	$\mathcal{C}$ (MFLOPS)
CELP	40	300
KELP (full search)	40	8
KELP (32–32 search)	< 42	0.5
KELP (binary search)	80	0.16

With the KELP structure, however, a new synthesis structure must be used. The KELP synthesis structure is shown in Fig. 5.

QUANTIZATION OF THE SHORT TERM PREDICTOR

The short term predictor ($p = 10$) was calculated using a window size of 80 samples and a frame size of 160 samples (200% overlap). Every other frame

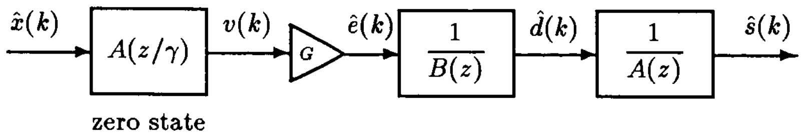

Fig. 5. The KELP Synthesis Structure

is quantized. Intermediate frames are found by linear interpolation between two adjacent frames.

The short term predictor is quantized based on the Line-Spectrum-Pairs [Kabal 1986]. Each pair is quantized in a two dimensional vector quantizer. The mean squared error was minimized during the design of the vector quantizers. Thirty-two bits were allocated for quantizing the short term predictor, with a {6,7,7,6,6} bit assignment. It was found that this assignment lead to a good perceptual quality. Furthermore, channel errors can be easily handled since the short term filter is quantized on a block by block basis with no inherent memory.

With 32 bits allocated to the short term model filter we require a bit rate of 1.6 kbps.

QUANTIZATION OF THE LONG TERM PREDICTOR

The long term predictor is optimized based on minimizing the open loop residual. It is updated every 10 msec.

The pitch predictor is chosen from a large codebook (128 levels) to minimize the open loop residual error. The codebook is designed using the Inverse Filter Matching Principle. The average residual error was minimized in the codebook design process. Thus the distortion measure is a simple matrix multiply, and the centroid calculation consists of averaging the covariance matrices in each cell [LeBlanc 1988]. The codebook elements were stabilized as discussed in [Ramachandran, 1987].

The pitch period is quantized to seven bits and is in the range (20,147). Thus, 1.4 kbps is allocated to quantizing the long term model filter.

GAIN QUANTIZATION

The gain is quantized on a log–mse scale to five bits. A Lloyd-Max scalar quantizer was used. The gain is updated every 5 msec. This leads to a rate of 1 kbps for the gain.

The resultant codec based on KELP (Fig. 4) has excellent speech quality. The total bit rate using the aforementioned quantization procedures is 6 kbps. By exploiting the intraframe memory of the various parameters, improved performance could be realized at a corresponding increase in complexity. Note, however, that the introduction of memory would lead to a degradation in the presence of channel errors.

CONCLUSION

This paper introduced a novel method to greatly reduce the complexity of the CELP algorithm while maintaining excellent speech quality. The structure was modified into a form amenable to fast search non-exhaustive algorithms. The computational complexity has been reduced to 0.5 MFLOPS with only a very slight increase in memory requirements.

BIBLIOGRAPHY

J.D. Markel and A.H. Gray Jr., Linear Prediction of Speech, Springer-Verlag, New York, New York, 1976.

J. Makhoul, S. Roucos, and H. Gish, "Vector Quantization in Speech Coding," Proceedings of the IEEE, Vol. 73, No. 11, November 1985.

M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates," IEEE International Conference on Acoustics Speech and Signal Processing, Tampa Florida, March 1985.

I.M. Trancoso and B.S. Atal, "Efficient Procedures for Finding the Optimum Innovation in Stochastic Coders," IEEE International Conference on Acoustics Speech and Signal Processing, Tokyo Japan, 1986.

G. Davidson and A. Gersho, "Complexity Reduction Methods for Vector Excitation Coding," IEEE International Conference on Acoustics Speech and Signal Processing, Tokyo Japan, 1986.

P. Kroon, R.J. Sluyter, and E.F. Deprettere, "A Low Complexity Regular Pulse Coding Scheme With a Reduced Transmission Delay," IEEE International Conference on Acoustics Speech and Signal Processing, Tokyo Japan, 1986.

J. Chen, G. Davidson, and K. Zeger, "Speech Coding for the Mobile Satellite Experiment", IEEE International Conference On Communications, Seattle, Washington, 1987.

P. Kabal and R. Ramachandran, "The Computation of Line Spectral Frequencies Using Chebyshev Polynomials", IEEE Transactions On Acoustics Speech and Signal Processing, Vol. ASSP–34, No. 6, December 1986.

R. Ramachandran and P. Kabal, "Stability and Performance Analysis of Pitch Filters in Speech Coders", IEEE Transactions On Acoustics Speech and Signal Processing, Vol. ASSP–35, No. 7, July 1987.

W. LeBlanc, An Advanced Speech Coder Based on a Rate-Distortion Theory Framework, Masters Thesis, To be published, Carleton University, Ottawa, Canada, 1988.