IMPLEMENTATION AND DISCUSSION OF THE VSELP CODER

Lei Peng and Andrew Perkis

Department of Electrical and Computer Engineering
The University of Wollongong
P.O. Box 1144
Wollongong, NSW 2500

ABSTRACT
This paper presents a coder implementation procedure for the Vector-Sum Excited Linear Predictive Coder (VSELP), operating at 7950 bits/second. The quality of the coder is evaluated and some pitfalls in the coder specifications are identified. The paper also provides methodologies for verifying the search for the self-excitation sequences and the residual codebook searches. For the LPC analysis, a performance comparison between the fixed point covariance algorithm (FLAT) and the standard Autocorrelation method (AUTO) is performed showing the superiority of the FLAT algorithm.

INTRODUCTION

The VSELP (EIA,1989) which is being proposed as the U.S Standard for their digital cellular mobile telephone system, is a variation of the Code Excited Linear Predictive Coder (CELP) (Atal & Schroeder,1985). The main difference occurs in the pitch extraction, and the fact that the excitation is built using two orthogonalized codebooks.

This paper presents the coding algorithm with main emphasis placed on obtaining the self excitation signal (Rose & Barnwel III,1986) from the past history of the coded excitation itself, as well as searching the two residual codebooks. In addition, some of the pitfalls in the specification will be identified.

Since computing the self excitation signals and searching the codebook excitations are two vital sections in the implementation of the coder, we will also discuss methodologies for verification of these algorithms.

In the area of filter coefficient estimation, a variety of techniques have been fully developed. The VSELP coder uses an efficient fixed point covariance lattice algorithm (FLAT). This technique involves building an optimum (that which minimizes the residual energy) inverse lattice filter stage by stage. The paper discusses the performance of the FLAT algorithm and compares it with the Autocorrelation method developed by Markel and Gray (Markel & Gray, 1976). Three measures are used in the performance comparison; error signal energy, inspection of the performance considering a typical voiced and unvoiced frame and the complexity.

THE VSELP ALGORITHM DESCRIPTION AND DISCUSSION OF THE SPECIFICATION

The VSELP encoder shown in Figure 1 employs a long term predictive filter to produce the self excitation sequences, and utilizes two codebooks to model the residual signal after spectral whitening, operating at 7950 bps. The excitation signal in the coder, being the sum of the self excitation signal and two codebook vectors, gives good quality synthetic speech.

The input signal is sampled at 8 kHz before being weighted by a perceptual noise weighting filter. The resulting signal which will be compared with the zero state response of A(z/r) the excitation signals, has to be subtracted from the zero input response of A(z/r) due to filter state match. The speech frame (160 samples) is divided into 4 subframes, with all the parameters being evaluated, quantized, and transmitted within a subframe. The VSELP analysis consists of three basic functions;1) filter coefficient estimation, 2) search for the self-excitation sequence, 3) codebook excitation search. The VSELP synthesis contains the same functions except with the weighting parameter, r, removed from the synthesis filter. The major characteristics of the coder are described in the following sections.

Filter coefficient estimation

The filter coefficients for the analysis and synthesis filters are calculated once per frame with no pre-emphasis. The analysis interval is centered with respect to the center of the fourth subframe using a

21.25 ms retangular window. The FLAT algorithm is used to estimate the reflection coefficients which are coded using 38 bit non-uniform scalar quantization with more bits allocated to the first reflection coefficient because of sensitivity. Direct form filter coefficients converted by reflection coefficients are linearly interpolated for each subframe. The interpolated subframe filter coefficients have to be converted to reflection coefficients to check for filter stability. If the resulting filter is unstable, then uninterpolated coefficients are used.

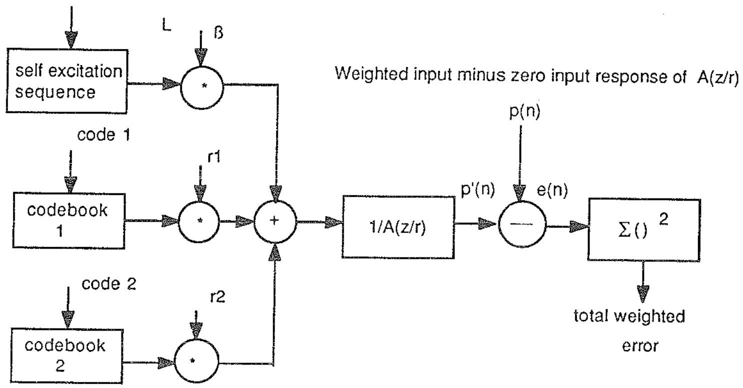


Figure 1. VSELP encoder

Search for the self-excitation sequence

The VSELP coder uses a closed loop approach to choose the self-excitation sequences in terms of lag which could be found by searching an adaptive codebook defined by the past output of the long term filter states.

The lag search in our coder is performed in every subframe whose value is varied from 40 to 167 instead of 20 to 147 as specfied in the VSELP standard, still, however representing the lag by 7 bits. The minimum squared prediction error (MSPE) criteria is applied. Vectors filtered by zero state response of A(z/r) are compared with weighted input minus the zero input response of A(z/r) for error weighting. The vector which minimizes the weighted error is the optimal vector, defined as the self-excitation sequence. The self-excitation gain is restricted to be positive, preserving the original sign of the chosen sequence.

Codebook search

Two codebooks are used in the VSELP coder, each consisting of M=7 predefined basis vectors. Codevectors which are coded with 7 bits are constructed as a linear combination of the M basis vectors. The codebook excitation is chosen from a codebook, containing 2^M codevectors, using the MSPE criteria.

The zero state response of each basis vector to A(z/r) must be computed for both codebooks, due to the error weighting between the codebook excitation and weighted input minus the zero input response of A(z/r). Vectors being selected are then correlated to each other. To dismantle the dependence an orthogonalization procedure is employed. Once the basis vectors have been filtered and orthogonalized, the two codebook excitation search procedures are identical.

Codebook excitation construction has such properties that a pair of ones complementary codevectors have equivalent values but opposite signs. Therefore, only half of the codevectors have to be evaluated. Also the codeword is sequenced using Binary Gray Code in which each successive codeword differs from the previous codeword in only one bit position so the current excitation can be efficiently computed from the previous one. The complexity of codebook excitation search can be significantly reduced by using these two properties.

Pitfall identification

During the simulation, some serious misprints and conceptual errors have been found in the VSELP specifications, which if not overcome, would result in the failure of coder implementation. These pitfalls are as follows:

1) page 2-12, line 339, equation (2.1.3.3.2.4-1);
2) page 2-21, line 720, equation (2.1.3.3.2.6.2-1).

The negative sign in the denominator in both equations should be positive as required by the filter coefficients, otherwise it would cause data being processed to end in floating point overflow error in simulation.

3) page 2-13, line 390-391: "if step 7 is done so that Fj(i,k), Bi(i-1,k-1), Cj(i,k-1), Cj(k,i-1) are updated together,"
given Fj-1(i,k)<>Bj-1(i,k)<>Cj-1(i,k)<>Cj-1(k,i)
so Fj(i,k)<>Bj(i-1,k-1)<>Cj(i,k-1)<>Cj(k,i-1)

It is seen that these four terms could not be updated together since they are not equal.

TESTING SCHEME

The correctness of the search for the self-excitation sequences and the codebook excitation search are of major concern in the success of coder implementation. It is therefore necessary to verify these algorithms in the coder simulation.

Testing scheme for verifying the self excitation sequences

Given the testing configuration in Figure 2, the residual signal output from the all-zero filter is used as the testing sequence. It splits into two branches; one serves as the past coded excitation signal being searched by the long term filter, another serves as the input signal being compared by the output of the long term filter. Since one frame consists of 4 subframes, it can be predicted that the lag obtained from the first subframe which produces the perfect self excitation sequence will be 0, as well as the second, third, and fourth subframes being 40, 80, 120 respectively. Since the self-excitation signal is the perfect match of the input signal, the pitch coefficient will be 1. Using this testing scheme in the coder, the algorithm implementation is correct if the results are as expected.

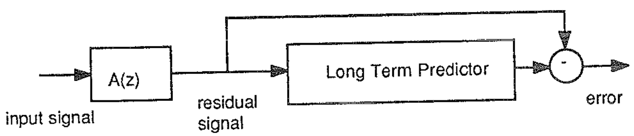


Figure 2. Self excitation search testing configuration

Testing scheme for verifying the codebook excitation search

The testing configuration is given in Figure 3. The residual signal is used to construct the test basis vectors shown in Figure 4 . For simplicity, 4 testing basis vectors are used which reduces complexity but still reflects the correctness of the algorithms accurately.

Combining 4 testing basic vectors will give the codebook excitation 16 entries. Each entry will be compared with the residual signal for error weighting. It is noticed that only when the entry of the codevector is 15, it will yield MSPE for codebook 1. Since 8 basic vectors are orthogonalized to each other, it is predicted that the codevector of the second codebook is also 15. The sum of gain factors for two codebooks will be 1. when this testing scheme is applied to the coder, the algorithm will be verified by achieving satisfactory results.

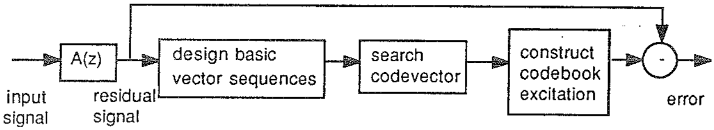


Figure 3. Codebook excitation search testing configuration

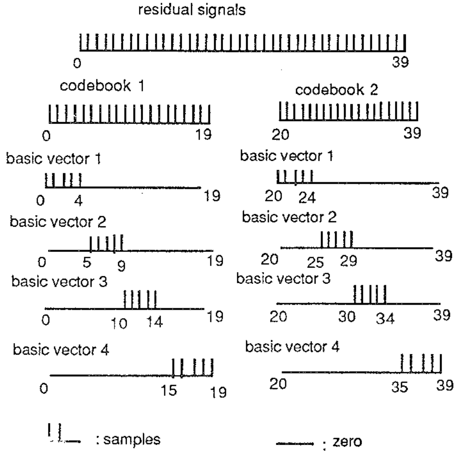


Figure 4. Testing basic vectors

COMPARISON STUDY OF "FLAT " AND "AUTOCORRELATION" ALGORITHM FOR LINEAR PREDICTION

The basic problem of linear prediction is to determine a set of filter coefficients directly from the speech signal in such a manner as to obtain a good estimate of the spectral properties of the speech signal. One approach to determine the coefficient is the Autocorrelation method (AUTO) proposed by Markel and Gray which views the technique as two steps, computing a matrix of correlation values and solving a set of linear equations. Another class of methods such as the FLAT algorithm used in our coder is called the lattice method, which has evolved as the action of combining the above two steps into a recursive algorithm. The major difference between the two methods is that in the lattice method the filter coefficients are obtained directly from the speech samples without an intermediate calculation of an autocorrelation function. The lattice method is also guaranteed to yield a stable filter without requiring the use of a window as it does in the AUTO method. The two methods are compared using the following three measures; error signal error, inspection of a typical voiced and unvoiced frames and complexity comparisons.

Error signal energy

The error signal obtained by filtering the input signals through an all-zero filter is a whitened version of the input signal in which the correlation between the signals have been removed. An utterance " glue the sheet to the dark blue background" is processed by an all-zero filter A(z) whose coefficients are predicted by the FLAT and the AUTO algorithms respectively. The error signal energy and mean value over this utterance are shown in Table1. It is indicated that results obtained by the FLAT algorithm yields a smaller value which gives a more flat spectrum than that of the AUTO algorithm. Thus the FLAT algorithm performs better.

Table 1. Error signal energy

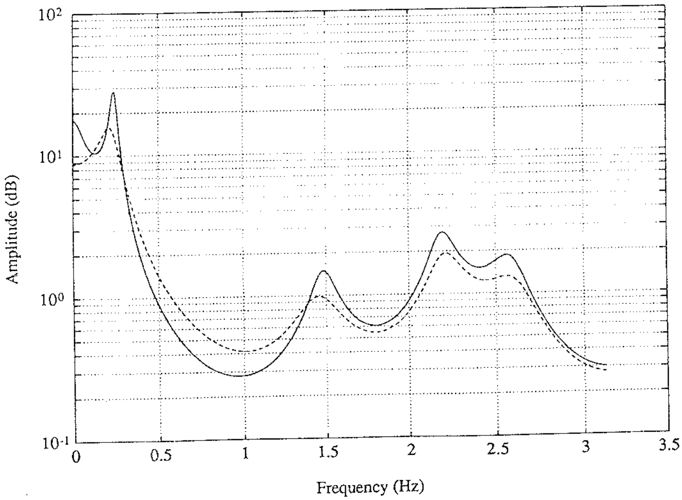
Algorithm	Squared-error dB	Mean dB
FLAT	-32.53933	0.001201
AUTO	-32.187824	0.002299

Table 2. Error signal energy for voiced and unvoiced frames

Frame	Algorithm	Squared-error (dB)
Voiced	FLAT	0.1741575
	AUTO	0.2553135
Unvoiced	FLAT	0.000787
	AUTO	0.001087

Typical voiced and unvoiced frame

A typical voiced frame (160 samples) and a typical unvoiced frame is chosen for performance inspection. The error signal energy for these two frames is shown in Table 2. It is shown that the FLAT algorithm always gives a smaller value which yields better results for both voiced and unvoiced sound.



——— : FLAT ------ : AUTO

Figure 5. Comparisons of voiced spectrums using the FLAT and the AUTO algorithms

Figure 5 presents the impulse response of the filter using the FLAT algorithm and the AUTO algorithm. It is seen that the spectrum obtained by the FLAT algorithm gives sharper peaks which contain more details compared with the AUTO algorithm. It shows that the FLAT algorithm indeed more accurately tracks the speech behavior.

Complexity comparison

A key factor in the choice of algorithm is its complexity. The degree of complexity should be as low as possible in order to reduce the computation time. Table 3 gives the complexity of estimating the coefficients for one frame using the FLAT and the AUTO algorithms.

Table 3. Complexity of the FLAT and the AUTO algorithms

FLAT	AUTO
0.26MIPS	0.25MIPS

MIPS-instruction/sec

It is seen that the complexity of the two algorithm is comparable.

In summary, the error signal energy obtained by the FLAT algorithm is approximately 0.3 dB less than that of the AUTO algorithm measured over an utterance, and the short term spectrum of a typical voiced frame shows more detail. Since the complexity of the two algorithms is comparable, it leads one to conclusion that the performance of the FLAT algorithm is superior to the AUTO algorithm for Linear Prediction analysis..

CONCLUSION

In this paper, we describe the algorithm for the VSELP coder implementation. The synthetic speech is evaluated using the segmental signal-to-noise ratio which gives 9.08 dB indicating that the VSELP coder produce good quality speech. To verify the search for the self-excitation signal and the codebook excitation searches, testing schemes have been designed accordingly. Also, this paper investigates the performance of the FLAT algorithm. This is done by comparing it with the AUTO algorithm through three measures. The results of comparing the two stress the prominent performance of the FLAT algorithm. The overall performance and qualities of the VSELP make it a promising candidate for lower bit rate applications.

REFERENCES

Atal, B.S. & Schroeder, M.R. (1985) " Stochastic Coding of Speech Signals At Very Low Bit Rates ", Proc. ICASSP.

EIA, Project number 2215 (1989) "Cellular System, Dual-Mode Subscriber Equipment Network, Equipment Compatibility Specification", IS-54 December .

Markel, J. D. & Gray A. H. Jr (1976) "Linear Prediction of Speech ", Springer Verlag.

Rose,R. C. & Barnwel III, T. P. (1987) " Quality comparison of Low Complexity 4800 bps Self Excitated and Code Excited Vocoders ", Proc. ICASSP.