A Robust CELP Coder with Source-Dependent Channel Coding

Rafid A. Sukkar (1) & **W. Bastiaan Kleijn** (2)
AT&T Bell Laboratories
(1) 200 Park Plaza, Naperville, IL 60566
Phone: (708) 713-5281
(2) 600 Mountain Ave, Murray Hill, NJ 07974

ABSTRACT

A CELP coder utilizing Source-Dependent Channel Encoding (SDCE) for optimal channel error protection is introduced. With SDCE, each of the CELP parameters are encoded by minimizing a perceptually meaningful error criterion under prevalent channel conditions. Unlike conventional channel coding schemes, SDCE allows for optimal balance between error detection and correction. Our experimental results show that our CELP system is robust under various channel bit error rates and displays a graceful degradation in SSNR as the channel error rate increases. This is a desirable property to have in a coder since the exact channel conditions cannot usually be specified *a priori.*

I. INTRODUCTION

Significant strides have been made in improving the speech quality of Code Excited Linear Prediction (CELP), making it a viable method for many telecommunication applications where bandwidth is scarce. In many of these applications, including mobile satellite communications, the speech coding algorithm must be robust in the presence of channel errors. CELP research efforts have focused mainly on improving the speech quality, and minimizing the computational complexity. Recently, more attention has been directed toward the robustness of the algorithm in the presence of channel errors [1].

In this paper a CELP system with source-dependent channel encoding scheme is introduced, extending earlier work described in [6]. For every CELP parameter, the source-dependent channel code is obtained by minimizing an appropriate distance measure. Compared to conventional forward error protection methods, SDCE is more efficient due to several factors. First, conventional error protection codes are designed without knowledge of the source coder implying that the bits that need to be protected must be hand picked, thereby providing only a rudimentary form of source-dependent channel coding. SDCE on the other hand provides error correction/detection such that highly probable quantization levels receive more accurate correction and/or serious errors are more likely to be detected. Second, with conventional methods, error correction/detection performance is predetermined, while with SDCE an optimal trade-off between error correction and detection is obtained. Third, conventional error correction codes are designed to perform exact error correction, with associated large increase in bit rate. With SDCE, significant improvement in performance can be obtained by reducing the impact of errors rather than reducing the number of errors. Also, error sensitivity can be reduced by an arbitrary amount using fractional bit allocation.

The organization of this paper is as follows. In the next section a brief description of our CELP system is given. In Section III SDCE is applied to each of the CELP parameters individually and performance with respect to channel errors is shown. Finally, in Section IV a complete CELP system with error protection bit allocation is given. Performance and experimental results are shown.

II. CELP CODER DESCRIPTION

The CELP system used here is based on the system described in [2]. Spectral information is transmitted as 10 line spectral frequencies and updated every 30 msec. Each 30 msec. frame is divided into four subframes for LPC excitation modeling. The LPC excitation modeling consists of two codebook searchs; an adaptive codebook search for modeling the speech periodicity, and a stochastic codebook search for modeling the speech randomness. The adaptive codebook has 128 overlapping entrees consisting of samples of previous frame excitations. The stochastic codebook is also overlapping, consisting of 512 entrees of center-clipped white Gaussian noise samples. However, only even numbered entrees are allowed for transmission implying a total of 256 codewords. A summary of the CELP parameter bit allocation without error protection is given in Table 1. The effective bit rate is 4233 bits/s.

parameter	bits/subframe	bits/frame
LSF(1)	N/A	4
LSF(2)	N/A	4
LSF(3)	N/A	4
LSF(4)	N/A	4
LSF(5)	N/A	4
LSF(6)	N/A	3
LSF(7)	N/A	3
LSF(8)	N/A	3
LSF(9)	N/A	3
LSF(10)	N/A	3
adap. bk index	7	28
adap. bk gain	4	16
stoch. bk index	8	32
stoch. bk gain	4	16
Total		127

Table 1. Bit allocation without error protection

III. SOURCE-DEPENDENT CHANNEL ENCODING

The first step in designing source-dependent channel codes is to define a suitable error criterion. For the CELP parameters an ideal error criterion would be a function of the final synthetic speech quality. However, because of the computational complexity, such an error criterion is unrealistic for the combinatorial optimization required to find good channel codes. Instead, any criterion that is monotonically related to the synthetic speech quality can be used to produce similar results.

Let r_j, $(j = 0,1,...J-1)$, be a quantized version of a given parameter, r. Let the available codewords be denoted as c_m, $(m = 0,1,...,M-1)$, where $M \geq J$. Our goal is to find an optimal mapping, $f(c_m)$ that maps the the codeword c_m into a quantization index j $(j = f(c_m))$. This optimal mapping is obtained by minimizing an appropriate error criterion. The error criterion takes on the following general form:

$$E = \sum_{m=0}^{M-1} P(c_m) \sum_{n=0}^{M-1} P(c_n \mid c_m)\, D(r_j, r_i) \tag{1}$$

where $P(c_m)$ is the a-priori probability that the codeword c_m is transmitted (this probability is zero for redundant codewords), and $P(c_n \mid c_m)$ is the transitional probabilities due to channel errors. The function $D(r_j,r_i)$ is a distance measure indicating the penalty for using r_i instead of r_j, where $j = f(c_m)$ and $i = f(c_n)$. The upper limit on the sum, M, is the total number of codewords.

If the error function, E, is evaluated as in Equation (1), the channel characteristics need to be defined. However, in many cases the channel characteristics are not well defined and stationarity cannot be guaranteed. Therefore, we like to modify the error criterion such that only broad assumptions are made about the channel, resulting in channel codes with performance that does not degrade significantly under varying channel conditions. A reasonable assumption to make is that the channel can alter at most a predefined number of bits in each codeword, where the most likely errors are weighted more heavily. For instance, single bits errors are weighted more heavily than double bit errors. For the derivation of channel codes in our CELP system, we assumed single bit errors only, though the method can easily be extended to cover any number of bit errors. We also assumed that all one bit errors are equally likely. The assumption of single bit errors is realistic if bit interleaving is employed and the channel performance is relatively good.

Based on the above assumptions the error criterion can now be written as

$$E = \sum_{m=0}^{M-1} P(c_m) \sum_{k=0}^{K-1} D(r_j, r_{j_k} \mid f(.)) \tag{2}$$

where, as before, $j = f(c_m)$, and j_k is the quantization index corresponding to the codeword c_m with bit k inverted $(j_k = f(c_{m_k}))$. The function $D(r_j, r_{j_k} \mid f(.))$ is the penalty function associated with replacing r_j with r_{j_k}, given a specific mapping function, f(.). Here, K is the total number of bits in each codeword. The error function, E, is minimized with respect to the mapping function, f(.). This minimization is highly non-linear requiring a simulated annealing-type procedure [3,4] to find the optimal f(.).

If redundant codewords are used, then the minimization of E can be used for both error correction and detection. For error detection, an additional fictitious level is introduced. Any of the redundant codewords can map into this fictitious quantization level. Receipt of a redundant codeword mapping into this level would indicate a transmission error, triggering the error recovery procedure. The penalty for synthesizing with this fictitious quantization level can be determined and must be used during optimization. Error correction is performed by assigning more than one codeword to map into a single quantization level. This error detection/correction SDCE scheme results in an optimal trade-off between error correction and detection. In our system the penalty function in Equation (2) depends on the CELP parameter at hand. We will now treat each parameter separately.

Line Spectral Frequencies

The penalty function for the line spectral frequencies is based on the cepstral distance measure [5]. For each LSF parameter the distance measure is defined as

$$D(LSF_j(p), LSF_{j_k}(p) \mid f(.)) = E[\mathbf{c_j}^T \mathbf{c_{j_k}} \mid f(.)] \qquad (3)$$

where p is the LSF number, $\mathbf{c}_j$ is the cepstral coefficient vector based on the quantized LSF's, and $\mathbf{c}_{j_k}$ is the cepstral coefficients vector corresponding to the the quantized LSF's with $LSF_{j_k}(p)$ replacing $LSF_j(p)$. This LSF replacement may, however, result in unrealistic LSF vectors since the monotonicity property may be lost. These cases can be thought of as error detect cases where the decoder receives an unrealistic LSF vector due to channel errors. Therefore, the strategy used in these cases should be the same as the strategy used in the decoder when unrealistic LSF vectors are received. If p is odd, then the previous frame $LSF_j(p)$ and $LSF_j(p+1)$ are substituted for the present frame $LSF_{j_k}(p)$ and $LSF_j(p+1)$, respectively. If p is even, then the previous frame $LSF_j(p)$ and $LSF_j(p-1)$ are substituted for the present frame $LSF_{j_k}(p)$ and $LSF_j(p-1)$, respectively. The monotonicity is checked again and if the resulting LSF vector is still unrealistic, then the whole LSF vector of the previous frame is used to compute the penalty function for the present frame. The expected value in Equation (3) is computed over all voiced frames in a database consisting of 24 sentences.

Parameter	NBC Mean Error	Gray Mean Error	SDCE Mean Error
LSF(1)	9.55	7.77	6.28
LSF(2)	13.62	11.16	10.41
LSF(3)	13.44	11.09	10.15
LSF(4)	13.73	10.73	10.66
LSF(5)	15.08	12.58	12.36
LSF(6)	16.89	15.26	14.65
LSF(7)	15.78	14.99	14.61
LSF(8)	13.00	11.09	10.29
LSF(9)	10.54	8.34	7.97
LSF(10)	7.92	7.27	7.07

Table 2. LSF error criterion after minimization

The error function of Equation (2) incorporating the penalty function of Equation (3) is minimized using a simulated annealing procedure [3]. With no redundant bits, the results of the minimization are given in Table 2. For comparison purposes the penalty function corresponding to the Natural Binary Code (NBC) and Gray code are also given.

Table 2 shows that SDCE consistently outperforms the other two schemes with a large improvement for LSF(1). The large improvement for LSF(1) is attributed to the fact that the the penalty functions associated with the quantization levels of LSF(1) have larger variation in dynamic range than the penalty functions of LSF(2)-LSF(10). This is typical of SDCE where serious errors are weighted more heavily than less serious errors in the optimization process.

To test SDCE on actual speech, the 24 sentence database was used to obtain a channel bit stream that was then corrupted on an LSF by LSF basis. For every other frame in the bit stream, one random bit of a given LSF codeword was inverted. The resulting Segmental Signal-to-Noise Ratio (SSNR) between the original speech and the synthetic speech over all voiced frames in the database is given in Table 3. The clear channel SSNR is 9.96 dB.

Parameter	NBC SSNR(dB)	Gray SSNR(dB)	SDCE SSNR(dB)
LSF(1)	6.58	6.96	7.35
LSF(2)	6.16	6.60	6.60
LSF(3)	7.36	7.71	7.72
LSF(4)	7.84	8.27	8.25
LSF(5)	9.01	9.15	9.20
LSF(6)	9.07	9.20	9.34
LSF(7)	9.35	9.46	9.45
LSF(8)	9.56	9.62	9.62
LSF(9)	9.80	9.83	9.82
LSF(10)	9.87	9.87	9.88

Table 3. SDCE actual speech performance after optimization

Again, Table 3 shows a significant improvement is obtained for LSF(1), while only marginal improvement to no improvement is obtained for LSF(2)-LSF(10).

To take better advantage of SDCE properties, we can consider combining two or more quantized LSF's and code them as one parameter (i.e., vector coding). The advantage of this is that by combining two or more quantized LSF's some combination of quantization levels become unrealistic due to the LSF monotonicity property. These levels, which can correspond to any fraction of a bit, can be used by the SDCE procedure as redundant levels for error correction and detection. In the scalar case, these unrealistic levels correspond to receiving an unrealistic LSF vector, thereby providing only a rudimentary form of error detection. In the vectorized case, SDCE uses these redundant levels to strike an optimal balance between error correction and detection.

Because of the computational complexity involved in the optimization process, we chose to combine only two LSF's at a time, although the coding efficiency increases as more LSF's are combined. The results for 0-bit redundancy are tabulated in Table 4. The SSNR column in Table 4 represents the average SSNR of the synthetic speech over voiced frames in the database after inverting a single bit in a combined LSF codeword every other frame. To compare with the performance of the scalar case, we have generated in Table 5 SSNR values for the case of channel encoding each LSF individually but corrupting, every other frame, a random bit taken from the set of bits spanning the codewords of two LSF's. The results of Tables 4 and 5 indicate that the combined case gives a significant improvement for LSF(1,2), without adding extra bits, or reducing the number of valid quantization levels. These results also demonstrate the ability of SDCE to use non-integer bit redundancy for error protection.

Parameter	Bits	Quant/Redun Levels	Mean Error	SSNR dB
LSF(1,2)	8	190/66	7.07	7.12
LSF(3,4)	8	190/66	9.01	7.83
LSF(5,6)	7	96/32	11.82	9.26
LSF(7,8)	6	52/12	10.72	9.60
LSF(9,10)	6	52/12	6.42	9.87

Table 4. SDCE performance of combined LSF's with 0-bit redundancy (vectorized case).

Parameter	Bits	Quant/Redun Levels	SSNR dB
LSF(1,2)	8	256/0	6.70
LSF(3,4)	8	256/0	7.76
LSF(5,6)	7	128/0	9.18
LSF(7,8)	6	64/0	9.55
LSF(9,10)	6	64/0	9.85

Table 5. SDCE performance of the scalar LSF optimization.

Table 6 shows the performance of the combined case after adding one-bit redundancy. Comparing Tables 4 and 6, it is clear that one redundant bit results in a significant improvement in SSNR for LSF(1,2) and LSF(3,4). These results indicate that the speech quality is susceptible to errors in LSF(1)-LSF(4) and is only marginally sensitive to errors in LSF(5)-LSF(10).

Parameter	Bits	Quant/Redun Levels	Mean Error	SSNR dB
LSF(1,2)	9	190/322	4.27	7.75
LSF(3,4)	9	190/322	5.46	8.52
LSF(5,6)	8	96/160	6.16	9.50
LSF(7,8)	7	52/76	3.50	9.80
LSF(9,10)	7	52/76	3.34	9.88

Table 6. SDCE performance of combined LSF's with 1-bit redundancy (vectorized case).

Codebook gain parameters

The penalty function used for the adaptive and stochastic gain parameters is derived from the error criteria used in CELP for choosing the codebook winning indices and determining the optimal gain. The penalty function is written as

$$D(\lambda_j, \lambda_{j_k} \mid f(.)) =$$

$$E\left[10\log\left[(\lambda_{j_k}\mathbf{s}_w-\mathbf{t})^T\mathbf{H}^T\mathbf{H}(\lambda_{j_k}\mathbf{s}_w-\mathbf{t})\right] \mid f(.)\right], \quad (4)$$

where λ_j is the optimal quantized gain, and λ_{j_k} corresponds to the quantization level obtained by inverting bit k of the codeword assigned to λ_j. The matrix $\mathbf{H}$ is the matrix which transforms the excitation vector of CELP into its zero-state response of the inverse linear predictive filter [2]. The vector $\mathbf{s}_w$ is the winning entree into the codebook, and $\mathbf{t}$ is the target excitation vector in CELP. The expected value is carried over voiced frames in the 24 sentence database.

Table 7 shows the adaptive codebook gain performance under 1-bit channel errors for various encoding schemes. The mean penalty is measured as a mean signal-to-noise ratio defined as, $E[10\log(\mathbf{t}^T\mathbf{t})] - D(\lambda_j, \lambda_{j_k} \mid f(.))$. The distribution of the adaptive gain quantization levels is highly non-uniform with values close to unity having the highest probability. The third row of Table 7 shows an example where a non-integer number of bits is used for protection. In this example the number of quantization levels is dropped from 16 to 12 by eliminating four quantization levels. With only 4 redundant levels a significant improvement is achieved at a minimal cost to the clear channel SSNR performance which dropped from 9.96 dB to 9.78 dB. This large improvement with only a small number of redundant levels is typical of SDCE and is the result of the channel code protecting the quantization levels with high probability only.

Adding a redundant bit results in a significantly higher performance. This performance is even higher than that obtained when 1-bit parity is used despite the fact that the parity bit was not subjected to bit errors. The error recovery strategy used in the error detect cases was to repeat the previous frame adaptive codebook gain.

Code	Bits	Quant/Redun Levels	Mean Penalty dB	SSNR dB
Gray	4	16/0	-0.89	2.38
SDCE	4	16/0	0.71	2.69
SDCE	4	12/4	3.59	6.54
Parity	5	16/16	3.91	7.15
SDCE	5	16/16	4.56	8.21
SDCE	6	16/48	4.95	9.37

Table 7. Adaptive codebook gain performance

The performance of the stochastic codebook gain displays the similar trends to those of the adaptive codebook gain, although the improvements over Gray code are not as dramatic. This is because of the smaller dynamic range of the stochastic codebook compared to the adaptive codebook, and the more uniform statistical distribution of the quantization levels. A complete discussion of the stochastic codebook gain performance is given in [6].

Codebook indices

The penalty function used here is similar to the one used for the gain parameters. It is defined as

$$D(s_j, s_{j_k} \mid f(.)) =$$

$$E\left[10\log\left[(\lambda_w s_{j_k} - t)^T H^T H(\lambda_w s_{j_k} - t)\right] \mid f(.)\right], \quad (5)$$

where $\mathbf{H}$ is define as before, and λ_w is the optimal quantized gain. The vector s_j is the winning codebook entree, and s_{j_k} correspond to the codebook entree obtained by inverting bit k of the codeword associated with s_j.

Table 8 shows the performance results of various methods of encoding the adaptive codebook index. When redundant levels were employed with SDCE, the error detection/correction optimization resulted in mostly error detection. The error recovery strategy used in the optimization was to repeat the previous frame adaptive codebook index. The SDCE performance is slightly lower than that of the parity-bit performance since in the latter procedure the parity bit was again assumed to be immune against channel errors. Additional redundant codewords result in some improvement in performance. SDCE does provide a significant advantage if only a small number of redundant codewords are available as the third row of Table 8 indicates. The associated decrease in clear-channel SSNR performance is minimal; from 9.96 to 9.85 dB.

Code	Bits	Delays	Redun Cdwds	Mean Penalty dB	SSNR dB
Gray	7	21-148	0	-0.21	2.25
SDCE	7	21-148	0	0.25	2.14
SDCE	7	21-118	30	1.49	3.14
SDCE	8	21-148	128	2.88	4.22
Parity	8	21-148	128	2.95	4.28
SDCE	9	21-148	384	3.19	4.62

Table 8. Adaptive codebook index performance

The behavior of the penalty function of the stochastic codebook does not show regularity similar to that of the adaptive codebook index. The only structure results from the overlapping nature of the stochastic codebook. The difference between clear-channel and 1-bit error performance is smaller than that of the adaptive codebook. However, SDCE gives a relatively large improvement over the Gray code since it can take advantage of the irregular structure of the penalty functions. A complete evaluation of the stochastic codebook index performance is given in [6].

IV. CELP WITH SDCE

The error protection bit allocation for the CELP parameters were based on the results of the previous section. Table 9 shows the total bit allocation for our CELP coder. The effective channel bit rate is 4800 bits/s. All of the parameters, regardless of the number of redundant levels used were channel encoded using SDCE. The line spectral frequencies were encoded as pairs as described in Section III. Most of the redundant bits were assigned to the adaptive codebook index and gain parameters since the synthetic speech quality is very sensitive to distortion in the speech periodicity during voiced regions. The rest of the redundant bits were assigned to the combined encoding of LSF(1) and LSF(2) (LSF(1,2)). The 24 sentence database was used to evaluate the overall coder performance by corrupting the associated CELP bit stream with errors at various rates. Table 10 displays the SSNR performance for this SDCE CELP coder computed over voiced frames in the database. For comparison, the performance of the basic 4233 bits/s coder (Table 1) using Gray code to channel encode the parameters is also shown in Table 10. The results of Table 10 show that, for SDCE CELP, there is a graceful degradation in performance as the error rate is increased from 0% to 1%. At error rates exceeding 1% the performance drops substantially because at such rates the probability of multiple bit errors per parameter is high. Since the SDCE optimization is carried over 1-bit errors, this substantial drop is expected. However, if multiple bit errors are likely, then the optimization process can be extended to cover such errors.

parameter	bits/subfrm (redun. levels)	bits/frm (redun. levels)
LSF(1,2)	N/A	9 (322)
LSF(3,4)	N/A	8 (66)
LSF(5,6)	N/A	7 (32)
LSF(7,8)	N/A	6 (12)
LSF(9,10)	N/A	6 (12)
adap. bk index	9 (384)	36
adap. bk gain	6 (48)	24
stoch. bk index	8 (0)	32
stoch. bk gain	4 (0)	16
Total		144

Table 9. CELP total bit allocation

Error Rate	Basic CELP SSNR (dB)	SDCE CELP SSNR (dB)
0%	9.96	9.96
0.1%	7.29	8.88
0.3%	5.45	7.26
0.5%	3.20	6.00
1.0%	1.42	4.10
2.0%	−0.35	1.69

Table 10. Overall CELP performance

V. CONCLUSIONS

A CELP coder utilizing source-dependent channel encoding was introduced. Unlike conventional error protection methods, SDCE allows for non-integer bit redundancy and strikes an optimal trade-off between error detection and correction. With SDCE, only broad assumptions need to be made about the channel providing, as our experimental results show, a graceful degradation in performance as the channel error rate increases. Although single bit errors were assumed throughout the paper, the extension to include multiple bit errors is straight forward. Also, more sophisticated error recovery strategies can be used in the error detect cases to further improve performance.

REFERENCES

1. **R. V. Cox, W. B. Kleijn, and P. Kroon.** 1989. Robust CELP coders for noisy backgrounds and noisy channels. In *Proceedings 1989 IEEE Int. Conf. Acoust. Speech Sig. Proc.* Vol. 2, pp. 739-742.

2. **W. B. Kleijn, D. J. Krasinski, and R. H. Ketchum.** 1988. An efficient stochastically excited linear predictive coding algorithm for high quality low bit rate transmission of speech. *Speech Communication.* Vol. 7, pp. 305-315.

3. **S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi.** 1983. Optimization by simulated annealing. *Science.* Vol. 220, pp. 671-680.

4. **A. E. Al Gamal, L. A. Hemachandra, I. Shperling, and V. K. Wei.** 1987. Using simulated annealing to design good codes. *IEEE Trans. Information Theory,* Vol. 33, pp. 116-123.

5. **A. H. Gray and J. D. Markel.** 1976. Distance measures for speech processing. *IEEE Trans. Acoust., Speech, and Sig. Proc.* Vol. 24, pp. 380-391.

6. **W. B. Kleijn.** 1990. Source-Dependent Channel Coding for CELP. In *Proceedings 1990 IEEE Int. Conf. Acoust. Speech Sig. Proc.* Vol. 1, pp. 1-4.