Source Optimized Channel Decoding and Unequal Error Protection of CELP Encoded Speech[*]

Ali Nazer and *Fady Alajaji*

Mathematics and Engineering
Department of Mathematics and Statistics
Queen's University
Kingston, Ontario, Canada K7L 3N6

Abstract

The reliable transmission of Federal Standard CELP 1016 encoded speech over very noisy communication channels is investigated. First, the interframe and intraframe redundancy present in the CELP 1016 parameters is quantified using second-order Markov chains. It is shown that over one-quarter of the CELP bits in every frame of speech are redundant. An unequal error protection (UEP) coding scheme, which exploits this residual redundancy, is next proposed for the transmission of the CELP parameters over BPSK-modulated AWGN and independent Rayleigh fading channels. It employs rate-compatible convolutional (RCPC) codes used in conjunction with maximum a posteriori (MAP) soft decision decoding. Experimental results indicate substantial coding gains (up to 7 dB in $\mathbf{E_b/N_0}$) over conventional systems that utilize equal error protection and maximum likelihood (ML) decoding.

1 Introduction

The role of the *source encoder* is to transform the input signal into a more compact form. Ideally all of the redundant bits are removed in the source compression phase. The *channel encoder* then adds a certain amount of controlled redundancy to the input signal. This redundancy – under the form of an *error-control code* – is used to protect the information against the effects of channel noise. Traditionally, source and channel coding have been treated as separate entities; this approach is known as *tandem source-channel coding.* This is justified by Shannon's Separation Principle [10], which states that the source and channel coding functions can be designed independently from each other without a loss in the optimality of the system. However, Shannon's findings were asymptotic in nature – assuming no limit on complexity or delay. Recently, systems with *jointly* designed source and channel coding operations have been shown to outperform tandem systems under practical limitations such as finite block lengths. In this work, we consider *joint source-channel coding* methods for the reliable communication of Federal Standard CELP 1016 encoded speech [6].

The paper is organized as follows. In Section 2, the redundancy in the CELP 1016 parameters is quantified via second-order Markov models. In Section 3, we propose a joint source-channel coding scheme for the transmission of the CELP parameters over very noisy AWGN and Rayleigh fading channels. This coding scheme employs unequal error protection through the use of RCPC codes, and source optimized channel decoding via MAP soft-decision detection. The overall system model is briefly described in Section 4, and experimental results are presented in Section 5. A summary is stated in Section 6.

2 CELP 1016 Residual Redundancy

One frame of Federal Standard CELP 1016 consists of 10 *Line Spectral Pair* (LSP) parameters which model the signal's short term spectrum. This coding technique also makes use of an *adaptive* and *stochastic* codebook, which simulate the human speech's voiced and unvoiced excitations, respectively. The adaptive codebook is represented through four *pitch delay* and four *pitch gain* parameters per frame. Similarly, the stochastic codebook has four *codebook gain* and four *index* parameters.

All the parameters are of different bit lengths. For consistency we chose to quantify the redundancy in

---

[*] This work was supported in part by NSERC.

the 3 most significant bits (MSB's) of each CELP parameter. Let the random process, $\{U_{i,j}\}$, represent the three most significant bits of the i^{th} quantized CELP parameter in frame j. Thus, $\mathbf{U}_j$ is a vector consisting of the quantized CELP coefficients (3 MSB's) in frame j. We assume that the process, $\{\mathbf{U}_j\}_{j=1}^{\infty}$, is block stationary.

To quantify the residual redundancy inherent in the CELP parameters, we model $\{\mathbf{U}_j\}_{j=1}^{\infty}$ as a 2^{nd} order Markov process. It assumes that $U_{i,j}$ is independent of all previous parameters conditioned on knowing the immediately preceding CELP parameter of the same type, $U_{i-1,j}$, and the corresponding CELP parameter in the previous frame, $U_{i,j-1}$.

The entropy rate of the process $\{\mathbf{U}_j\}_{j=1}^{\infty}$ is given by

$$H(\mathcal{U}) = \lim_{n \to \infty} H(\mathbf{U}_n | \mathbf{U}_{n-1}, \mathbf{U}_{n-2}, .., \mathbf{U}_1). \quad (1)$$

$H(\mathcal{U})$ represents the minimum number of bits per frame required to describe $\{\mathbf{U}_j\}$. Thus, the total residual redundancy (per frame), ρ_T, of $\{\mathbf{U}_j\}$ is

$$\rho_T \stackrel{\triangle}{=} log_2|\mathcal{U}| \; - \; H(\mathcal{U}), \quad (2)$$

where $|\mathcal{U}|$ is the size of the source alphabet, $\mathcal{U}$. The total redundancy, ρ_T, can be divided into two parts – the redundancy due to the non-uniformity of the source and the redundancy due to the source memory, ρ_D and ρ_M respectively:

$$\rho_T = \rho_D + \rho_M, \quad (3)$$

where

$$\rho_D \stackrel{\triangle}{=} log_2|\mathcal{U}| \; - H(U_1), \quad (4)$$

$$\rho_M \stackrel{\triangle}{=} H(U_1) - H(\mathcal{U}), \quad (5)$$

and

$$H(U_1) = - \sum_{u_1} Pr(U_1 = u_1) \cdot \log_2 Pr(U_1 = u_1). \quad (6)$$

A large training sequence (83,826 frames) from the TIMIT speech database [7] was applied to the Federal Standard CELP 1016 vocoder. For every frame of speech, CELP analysis was performed to arrive at 26 quantized CELP parameters. The relative frequency of transitions between the values of 3 MSB's of each codebook parameter were compiled to compute its Markov transition probabilities. These probabilities were used in equations (4) and (5) to compute ρ_D and ρ_M, respectively. The results are compiled in Table 1 where the values of ρ_D, ρ_M and ρ_T are provided for each CELP parameter as well as for the entire frame. Note that around 12.5 bits of the 30 high-order bits of the LSP parameters are redundant. If we calculate the total frame redundancy, we obtain that among the 78 high-order bits of the CELP parameters, 21 bits (or $\approx 27\%$) of them are redundant.

CELP Parameter	Redundancy		
	ρ_D	ρ_M	ρ_T
LSP	5.2747	7.2105	12.4852
Codebook Gain	4.0478	1.2544	5.3022
Pitch Gain	0.1832	1.4910	1.6742
Pitch Delay	0.7064	0.8266	1.5330
Codebook Index	0.0323	0.0321	0.0644
Total Frame	10.2444	10.8146	21.0590

Table 1: CELP 1016 Redundancy (in Bits/Frame).

3 Joint Source Channel Coding

3.1 Unequal Error Protection

In addition to being redundant, the CELP 1016 quantized parameters contribute differently to the reconstruction of the speech [9]. We employ unequal error protection (UEP) to allow various levels of protection for different parameters. Our UEP system consists of a family of punctured convolutional codes [2], known as rate compatible punctured convolutional (RCPC) codes [3].

Punctured convolution codes were introduced to attain higher rate $R = k/n$ convolutional codes from lower rate $R = 1/n$ codes. They can be attained by periodically perforating the output of low-rate convolutional codes (or mother codes), through a puncturing matrix. More specifically, a rate $P/(P+\delta)$ punctured convolutional code can be attained by periodically puncturing a rate $1/n$ mother code with a puncturing matrix, $\mathbf{A}(\delta)$, and a period P , where $\mathbf{A}(\delta)$ is an (n x P) matrix, and $\delta \in [1, (n-1)P]$ [4]. The puncturing matrices simply contain 0's, which specify the punctured (or not transmitted) output bits, and 1's, which specify the unpunctured bits.

Rate-compatible punctured convolutional (RCPC) codes are a sub-class of punctured codes [3]. The *rate compatibility restriction* simply states that all the code bits of a high rate punctured code must be used by all the corresponding lower rate codes in the same family. In other words, the puncturing matrix for the lower rate code, $\mathbf{A}(\epsilon)$ contains all the 1's of the puncturing matrix for the higher rate code, $\mathbf{A}(\delta)$. The above condition guarantees that no loss of distance (d_{free} of the code) occurs between the higher rate code and the lower rate code in a transitional phase [3].

RCPC codes can easily be applied to a UEP scheme, by ordering the information by importance, and applying lower rate codes to the more important bits and higher rate codes to the less important ones.

Decoding of RCPC codes, as well as regular punctured codes, is based only on the trellis of the mother code where the metric corresponding to the punctured bits is replaced by zero. Thus, a family of RCPC codes, corresponding to one period (P), can be decoded with the same trellis, as long as the different rates, δ, their corresponding puncturing matrices, $\mathbf{A}(\delta)$, and the bits they protect are known at the decoder.

3.2 MAP Soft Decision Decoding

We assume that the CELP parameters are channel encoded and sent over a memoryless channel. At the receiver, we consider a MAP soft-decision decoder that exploits the CELP residual redundancy [1] in combating channel noise. This decoder, which is based on a modified version of the Viterbi algorithm, chooses the code sequence $\hat{\mathbf{x}}^K = (\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_K)$ that minimizes

$$Pr(\mathbf{y}^K \mid \hat{\mathbf{x}}^K)\ Pr(\hat{\mathbf{x}}^K), \tag{7}$$

where $\hat{\mathbf{y}}^K = (\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_K)$ is the received sequence of length K.

We only consider AWGN and fully interleaved Rayleigh Fading channels. Thus, the above metric reduces to choosing $\hat{\mathbf{x}}^K$ that minimizes

$$\sum_{k=1}^{K} \parallel \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \parallel^2 \ - \ N_0 \ ln \ Pr(\hat{\mathbf{x}}^K), \tag{8}$$

where $\mathbf{a}_K$ is the sequence of Rayleigh fading coefficients (for AWGN, $\mathbf{a}_k$ is the all-one vector for all k), which we assume to be available at the decoder. The prior distribution $Pr(\hat{\mathbf{x}}^K)$ is estimated using the Markov model of the previous section in conjunction with a large training sequence from the TIMIT database [7].

4 System Model

The diagram of the overall system proposed for UEP channel coding of the CELP parameters is shown in Figure 1. The first step is the CELP coder which inputs a speech signal and outputs the CELP parameters: 10 LSP's, 4 Pitch Gains, 4 Pitch Delays, 4 Codebook Gains, and 4 Codebook Indices.

The next step consists of the channel encoder. We consider three different systems: uncoded, equal error protection (EEP) using a 32-state rate 3/4 convolutional code [5], and a 32-state base rate 1/3 RCPC code with period $p = 8$ [4]. In the EEP and UEP models, only 78 bits of a total 144 bits per CELP frame are convolutionally encoded:

- the 3 MSB's of all the 10 LSP parameters,
- the 6 MSB's of all the 4 pitch delay parameters,
- the 3 MSB's of all the 4 codebook gain parameters,
- and the 3 MSB's of all the 4 pitch gain parameters.

Remark that the 2^{nd} three MSB's of the pitch delays are coded, because of their important role in speech excitation, but they have not been modeled for their redundancy. Thus, we will decode them using the traditional Viterbi decoding algorithm. The remaining 66 bits are sent uncoded and hard decision decoded for all three transmission schemes.

The UEP scheme allows various levels of protection for the different CELP parameters. The RCPC coding rates are described in Figure 2. They are chosen based on the CELP parameters sensitivity study in [9]. Note that the 3 MSB's of the pitch gain parameters are sent uncoded. However, these parameters were modeled for their residual redundancy. Thus, the MAP soft decision algorithm can still be applied in their decoding phase.

Parameter	Code Rate
LSP 1-10	8/24
Pitch Delay 1 & 3	8/22
Pitch Delay 2 & 4	8/20
Codebook Gain 1-4	8/18
Pitch Gain 1-4	Uncoded

Table 2: Overall CELP Unequal Error Protection Scheme Using Mother Rate-1/3 Family of RCPC Codes.

The next block in Figure 1 is the BPSK modulation, followed by the channel. In this simulation two channels were used - the AWGN and the fully interleaved Rayleigh channel, where it is assumed that channel state information (CSI) is available at the decoding phase.

Next, MAP soft decision decoding is performed. Note that a modified Viterbi algorithm was used to decode the UEP system. This is because the redundancy was modeled three bits at a time, whereas the

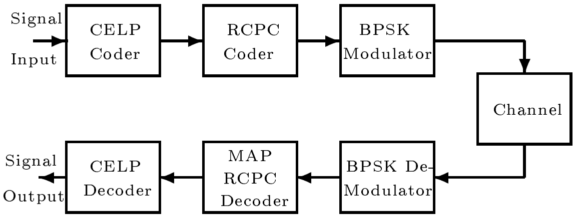

Figure 1: Block Diagram of the Communication System.

mother code is rate 1/3 in which the output is decoded one bit at a time. In order to use the MAP modified Viterbi metric, a special method is used as explained in [8]. Also, since the LSP parameters are the only ones with an ordering property, they are the only CELP parameters that undergo re-ordering after they are decoded. The final step is the synthesis of the speech from the corrupted CELP parameters.

5 Experimental Results

A large training sequence ($\approx$ 42 minutes) of speech was used from the TIMIT data base to estimate the prior CELP distributions needed for the MAP decoder. The testing sequence consisted of a 4753-frame (2.2 minutes) TIMIT speech sequence, half uttered by females and half uttered by males, with no speaker appearing in both the training and testing sequence. The performance criteria used are:

- The average speech distortion measure (see [8]), which is an average of seven different speech distortion measures of two different groups – cepstral measures and cosh measures. Note that this measure is averaged over all CELP subframes where subframes with either zero signal or noise energy are excluded from the average. When the channel is noiseless, the average speech distortion is 4.79 dB.
- Subjective listening tests that make pairwise comparisons between the different coding schemes.

The simulations are performed using a practical decoding delay of one frame in length (30 ms).

Figures 2 and 3 show the performance of the UEP MAP system, as well as two EEP schemes with MAP and with regular maximum likelihood (ML) decoding, respectively. The results for the uncoded system are also presented. Tables 3 and 4 show the coding gains first between EEP rate 3/4 with and without MAP, and second between UEP and EEP both with MAP decoding. All results are displayed for different values of E_b/N_0, where E_b is the energy per *information* bit[1], and N_0 is the one-sided power spectral density of the additive Gaussian noise.

At an average speech distortion of 6.0 dB, the UEP system with MAP outperforms the traditional EEP system with ML decoding by 3.2 dB and 5.0 dB, for the AWGN and Rayleigh fading channels respectively.

Our subjective listening tests consisted of pairwise comparisons between the two coded transmission systems at two different E_b/N_0's. Four different speech segments and fifty listeners – 25 male and 25 female – were tested. Before being tested the uncorrupted CELP encoded speech segments were played for each listener to "anchor" their perspective. The listeners were asked to choose the better sounding segment, or neither if they could not perceive a difference. Both systems used MAP decoding, so it was only a comparison of the effect of unequal error protection on the quality of speech reconstruction. The results showed that the UEP mother rate 1/3 system clearly performed better than the EEP 3/4 system [8]. For a demonstration of the results refer the following internet site: *http://markov.mast.queensu.ca/~nazera/*.

6 Summary

We investigated the problem of the reliable transmission of CELP 1016 speech parameters over very noisy BPSK-modulated AWGN and Rayleigh fading channels. A second-order Markov model was proposed to generate the CELP parameters and to quantify the amount of residual redundancy they exhibit. It was shown that over one-quarter of CELP bits in every frame of speech were redundant. We next proposed and implemented a joint source-channel coding scheme that employs: (i) UEP via a family of RCPC codes to provide additional protection for the important CELP parameters; and (ii) Optimal MAP soft-decision detection that utilizes the CELP residual redundancy in combating channel noise. Experimental results indicate that the proposed UEP-MAP scheme is significantly robust particularly during severe channel conditions; it also offers considerable performance improvements over traditional EEP systems and systems that employ ML decoding.

References

[1] F. I. Alajaji, N. Phamdo, and T. E. Fuja, "Channel Codes that Exploit the Residual Redundancy in CELP-Encoded Speech," *IEEE Trans. on Speech and Audio Proc.*, vol. 4, no. 5, pp. 325-336, Sept. 1996.

---

[1] Recall that $E_b = \frac{E_s}{R}$, where R is the channel code rate and E_s is the energy per *channel symbol.*

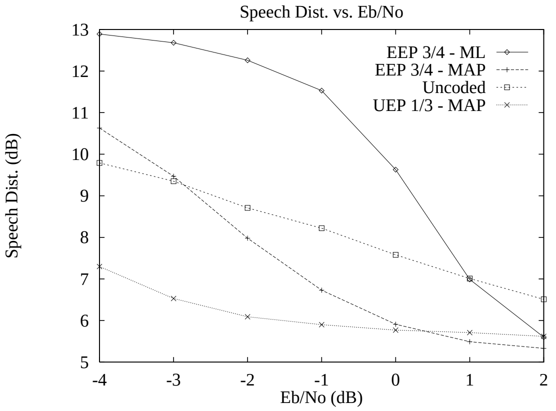

Figure 2: Average Speech Distortion for Different Coding Schemes of CELP Parameters over the AWGN Channel.

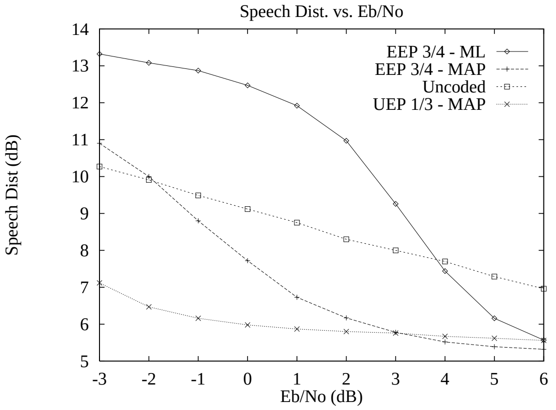

Figure 3: Average Speech Distortion for Different Coding Schemes of CELP Parameters over the Rayleigh Fading Channel.

Coding Gains	Avg. Speech Dist. (dB)		
	6.0	6.5	7.0
EEP 3/4 – MAP vs. EEP 3/4 – ML	+1.82	+2.07	+2.21
UEP 1/3 – MAP vs. EEP 3/4 – MAP	+1.42	+2.21	+2.39
Total Gain UEP 1/3 – MAP vs. EEP 3/4 – ML	+3.24	+4.28	+4.61

Table 3: Coding Gains for Base Rate 1/3 Unequal Error Protection over Rate 3/4 Equal Error Protection for the AWGN Channel.

Coding Gains	Avg. Speech Dist. (dB)		
	6.0	6.5	7.0
EEP 3/4 – MAP vs. EEP 3/4 – ML	+2.47	+3.20	+3.59
UEP 1/3 – MAP vs. EEP 3/4 – MAP	+2.55	+3.46	+3.54
Total Gain UEP 1/3 – MAP vs. EEP 3/4 – ML	+5.02	+6.66	+7.14

Table 4: Coding Gains for Base Rate 1/3 Unequal Error Protection over Rate 3/4 Equal Error Protection for the Rayleigh Fading Channel.

[2] J.B. Cain, G.C. Clark, and J.M. Geist, "Punctured convolutional codes of rate $(n-1)/n$ and simplified maximum likelihood decoding," *IEEE Trans. on Info. Theory*, vol. IT-25, pp. 97-100, Jan. 1979.

[3] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications," *IEEE Trans. on Comm.*, vol. 36, no. 4, pp. 389-400, April 1988.

[4] L.H.C. Lee, "New Rate-Compatible Punctured Convolutional Codes for Viterbi Decoding," *IEEE Trans. on Comm.*, vol. 42, no. 12, pp. 3073-3079, Dec. 1994.

[5] S. Lin and D.J. Costello, Jr., *Error Control Coding*, Englewood Cliffs, NJ: Prentice Hall, 1983.

[6] National Communications System (NCS) "Details to Assist in implementation of Federal Standard 1016 CELP," Office of the Manager, NCS, Arlington, VA, Tech. Inform. Bulletin 92-1.

[7] National Institute of Standards and Technology (NIST), "The DARPA TIMIT acoustic-phonetic continuous speech Corpus CD-ROM," NIST, Oct. 1990.

[8] A. Nazer, "Joint Source Channel Coding and Unequal Error Protection of CELP 1016 Encoded Speech over Very Noisy Channels," *M.Sc. Thesis*, Queen's University, 1998.

[9] D. E. Ray, "Reed-Solomon Coding for CELP Error Detection and Control in Land Mobile Radio," *M.Sc. Thesis*, University of Maryland, 1993.

[10] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Sys. Tech. Journal*, vol. 27, pp. 379-423, July 1948.