Unequal error protection and source-channel decoding of CELP speech

Ali Nazer and *Fady Alajaji*

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

E-mail: `fady@mast.queensu.ca`

Abstract: The reliable communication of FS CELP 1016 encoded speech over very noisy channels is investigated. 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 coding scheme, which exploits this residual redundancy, is proposed for sending the CELP parameters over Gaussian and Rayleigh fading channels. Simulations indicate substantial coding gains over conventional systems.

To appear in *IEE Electronics Letters*, Vol. 38, No. 7, 28th March 2002.

Introduction: During the last decade, several works have shown the practical advantages of *joint source-channel coding* (JSSC) systems, where the source and channel codes are jointly designed, over tandem systems where the two codes are treated separately. We consider a JSSC method for the robust communication of FS CELP 1016 encoded speech [1]. We propose an unequal error protection (UEP) and source-optimized channel coding scheme for the transmission of all the CELP parameters over additive white Gaussian noise (AWGN) and Rayleigh fading channels. This extends the work in [2], where equal error protection (EEP) coding focusing only on the CELP Line Spectral Pair (LSP) coefficients was presented. Our method exploits the residual redundancy within all the CELP parameters (including the LSP's) and employs rate-compatible punctured convolutional (RCPC) codes under maximum a posteriori (MAP) decoding. This study can also be applied to other recently developed low rate vocoders such as MELP and others.

CELP 1016 Redundancy: CELP 1016 is a frame oriented vocoder that samples the input at 8kHz with each frame (of 30ms duration) producing 144 bits at a rate of 4.8 kbs [1]. We examine the redundancy in the three most significant bits (MSB's) of *each set* of CELP parameters: LSP's, pitch gains, pitch delays, codebook gains and indices. Each parameter set is modeled by a 2^{nd}-order block stationary Markov process to capture both inter and intra frame redundancies. A large training sequence (83,826 frames) from the TIMIT speech database [3] is applied to the vocoder to estimate the Markov probabilities of each parameter set and compute its entropy rate, its redundancy ρ_D due to its non-uniformity, its redundancy ρ_M due to its memory [2] and its total redundancy ρ_T: $\rho_T = \rho_D + \rho_M$. The results are compiled in Table 1 where ρ_D, ρ_M and ρ_T are provided for each parameter and for the entire frame. We interestingly remark that of the considered 78 high-order CELP bits, more than 21 bits (or $\approx 27\%$) are redundant.

System Structure: In addition to being redundant, the CELP parameters contribute differently to the speech reconstruction. We hence apply UEP to allow various levels of protection for different parameters. Our system consists of a family of RCPC codes [4,5]. This is achieved by ordering the information by importance, and applying lower rate codes to more important bits and higher rate codes to less important ones. At the receiver, MAP soft-decision decoding that exploits their residual redundancy is employed. This decoder is based on the Viterbi algorithm and chooses the code sequence $\hat{\mathbf{x}}^K = (\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_K)$ that minimizes $Pr(\mathbf{y}^K \mid \hat{\mathbf{x}}^K)\ Pr(\hat{\mathbf{x}}^K)$, where $\mathbf{y}^K = (\mathbf{y}_1, \ldots, \mathbf{y}_K)$ is the received sequence of length K, which is the number of sent CELP parameters. We consider binary phase-shift keying modulated AWGN and fully interleaved Rayleigh fading channels with noise variance $N_0/2$. Thus, the above minimization metric reduces to

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

where $\mathbf{a}_K$ denotes the Rayleigh coefficients (for AWGN, $\mathbf{a}_k$ is the all-one vector) which we assume to be available at the decoder. The prior distribution $Pr(\hat{\mathbf{x}}^K)$ is estimated using the Markov model discussed earlier on the TIMIT training sequence.

We implement three coding systems: uncoded, EEP using a 32-state rate 3/4 convolutional code as in [2] and UEP using a 32-state base rate 1/3 RCPC code with period $p = 8$ [5,6]. In the EEP and UEP schemes, only 78 bits per 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. The remaining 60 bits are sent uncoded and are hard decision decoded for all transmission schemes. The RCPC coding rates, chosen based on the CELP parameters sensitivity, are described in Table 2.

A TIMIT testing sequence consisting of 4753 frames was used in our simulations which were performed using a practical decoding delay of one frame. The performance criterion used is the average speech distortion measure, which is an average of seven different speech distortion measures of two different types - cepstral and cosh measures [7]. Note that the minimum possible speech distortion (for a noiseless channel) is 4.79 dB.

The rate of the UEP scheme is $\frac{138}{252}$, while that of the EEP scheme of [2] is $\frac{138}{162}$. The performance of the two schemes with maximum likelihood (ML) and MAP decoding, as well as that of the uncoded scheme are shown in Figs. 1 and 2 for various E_b/N_0, where E_b is the energy per information bit. We clearly observe that the UEP-MAP scheme provides the best performance, particularly during severe and medium channel conditions. At a speech distortion of 6.0 dB, the gains for UEP-MAP versus EEP-MAP are 1.42 dB in Fig. 1 and 2.55 dB in Fig. 2; also, the gains for exploiting the CELP redundancy of EEP-MAP over EEP-ML are 1.82 dB and 2.47 dB in Figs. 1 and 2, respectively. Additional simulations using other performance criteria (such as the symbol error rate) and subjective listening tests are available in [6]. A listening demonstration can be accessed at the Internet site *http://markov.mast.queensu.ca/~nazera/*.

Acknowledgment: This work was supported in part by NSERC of Canada.

References

1 National Communications System (NCS): 'Details to Assist in implementation of Federal Standard 1016 CELP,' Office of the Manager, NCS, Arlington, VA, Tech. Info. Bull. 92-1, 1992.

2 ALAJAJI F., PHAMDO N. and FUJA T.: 'Channel codes that exploit the residual redundancy in CELP-encoded speech,' *IEEE Trans. Speech Audio Process.*, 1996, **4**, (5), pp. 325-336.

3 National Institute of Standards and Technology (NIST): 'The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM,' NIST, 1990.

4 HAGENAUER J.: 'Rate-compatible punctured convolutional codes (RCPC codes) and their applications,' *IEEE Trans. Commun.*, 1988, **36**, (4), pp. 389-400.

5 LEE L. H. C.: 'New rate-compatible punctured convolutional codes for Viterbi decoding,' *IEEE Trans. Commun.*, 1994, **42**, (12), pp. 3073-3079.

6 NAZER A.: 'Joint source channel coding and unequal error protection of CELP 1016 encoded speech over very noisy channels,' *M.Sc. Thesis*, Queen's University, 1997.

7 GRAY A. H. and MARKEL J. D.: 'Distance measures for speech processing,' *IEEE Trans. Acoustics, Speech Signal Process.*, 1976, **24**, (8), pp. 380-391.

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).

CELP 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: UEP scheme using mother rate-1/3 family of RCPC codes.

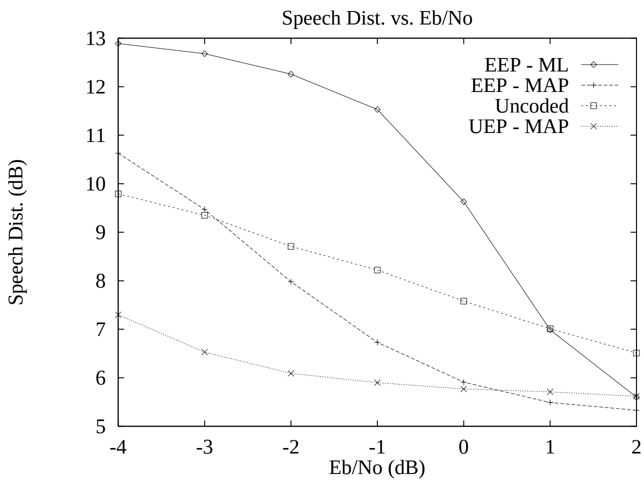

Figure 1: Average speech distortion for the AWGN channel.

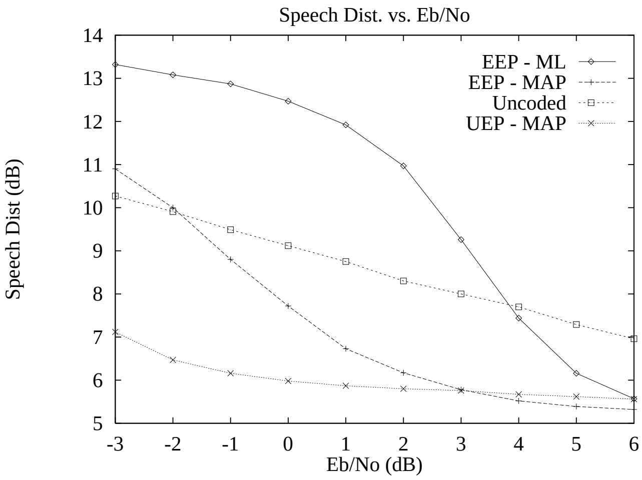

Figure 2: Average speech distortion for the Rayleigh channel.