Mathematics and Engineering Communications Laboratory

Technical Report

Unequal Error Protection and Source-Channel Decoding of CELP Speech over Very Noisy Channels

A. Nazer and *F. Alajaji*

March 1999

Unequal Error Protection and Source-Channel Decoding of CELP Speech over Very Noisy Channels

Ali Nazer and *Fady Alajaji*

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

Abstract

The reliable transmission of Federal Standard CELP 1016 encoded speech over very noisy communication channels is investigated. First, the inter-frame and intra-frame redundancies present in the CELP 1016 parameters are quantified via first- and 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 binary phase-shift keying (BPSK) modulated additive white Gaussian noise (AWGN) and independent Rayleigh fading channels. It employs rate-compatible convolutional (RCPC) codes used in conjunction with maximum a posteriori (MAP) soft-decision sequential decoding. Experimental results indicate substantial coding gains over uncoded systems and over conventional systems that utilize equal error protection and maximum likelihood (ML) decoding.

Index Terms: CELP, joint source-channel coding, unequal error protection, residual redundancy, RCPC codes, soft-decision MAP decoding, AWGN/Rayleigh Fading Channels.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant OGP0183645. Parts of this paper were presented at the Biennial Symposium on Communications, Kingston, Canada, June 1998.

1 Introduction

The role of the source code 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 code 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, resulting in what is known as a *tandem source-channel coding* system. This is justified by Shannon's Separation Principle [24], 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 are asymptotic in nature – assuming no constraints 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 (e.g., [1]-[5], [9], [10], [13], [15], [21], [25], [26]).

In this work, we consider *joint source-channel coding* methods for the robust communication of Federal Standard CELP 1016 encoded speech [6, 18]. More specifically, we propose and implement unequal error protection (UEP) and source-optimized channel coding schemes for the reliable transmission of all the CELP parameters over very noisy binary phase-shift keying (BPSK) modulated additive white Gaussian (AWGN) and independent Rayleigh fading channels. This work extends previous work in [2], where equal error protection (EEP) coding schemes using convolutional and Reed-Solomon codes were presented for the transmission of the Line Spectral Pair (LSP) parameters of the CELP encoded speech. These EEP schemes were used in conjunction with maximum a posteriori (MAP) soft-decision sequential decoding, thus exploiting the residual redundancy inherent in the LSP parameters. Our proposed methods, which exploit the residual redundancy within all the CELP parameters (including the LSP's), employ rate-compatible convolutional (RCPC) codes used in conjunction with MAP soft-decision decoding. Objective and subjective tests demonstrate considerable performance improvements over the results in [2] and systems that employ EEP and maximum likelihood (ML) decoding, particularly for severe channel conditions.

The paper is organized as follows. In Section 2, the intra-frame redundancy in the CELP 1016 parameters is quantified via first-order Markov models. Likewise, the intra-frame and inter-frame redundancies in the parameters are quantified using second-order Markov models. In Section 3, we present RCPC-based UEP schemes which employ source optimized channel decoding via MAP soft-decision detection. Two overall system models are introduced in Section 4, one for the transmission of the Line Spectral Pair parameters, and another for all the CELP parameters. In Section 5, quantitative experimental results are presented for both of these scenarios, and listening test results for the overall system. Finally, a summary is stated in Section 6.

2 CELP 1016 Residual Redundancy

Federal Standard CELP 1016 [6, 18] is a frame oriented vocoder that samples the input at 8kHz and breaks the corresponding samples into blocks, which are then processed. Each frame, containing 240 samples, is 30ms in duration and produces 144 bits (cf Table 1). The CELP 1016 frame is further subdivided into four 7.5ms sub-frames. The overall output rate is 4800 bits per second.

One frame of CELP 1016 consists of 10 *Line Spectral Pair* (LSP) parameters which model the signal's short term spectrum. CELP coding also makes use of *adaptive* and *stochastic* codebooks, 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 (or one per sub-frame). Similarly, the stochastic codebook has four *codebook gain* and four *index* parameters. In addition, there are some Hamming code bits, a synchronization bit, and an unused bit. The bit allocations for each set of parameters are presented in Table 1.

We examine the redundancy in the three most significant bits (MSB's) of *each set* of CELP parameters: the LSP's, the pitch gains, the pitch delays, the (stochastic) codebook gains and indices. For each set of parameters, let the random process, $\{U_{i,j}\}$, represent the three most significant bits of the i^{th} (quantized) CELP parameter in frame j, and let $\mathbf{U}_j = [U_{1,j}, U_{2,j}, \cdots, U_{l,j}]$, where l denotes the number of parameters per frame[1]. We assume that the process, $\{\mathbf{U}_j\}_{j=1}^{\infty}$, is block stationary.

In [2], two Markov models for $\{\mathbf{U}_j\}$ were introduced to estimate its entropy rate and thus compute the residual redundancy exhibited by the LSP parameters. We herein employ the same Markov models for $\{\mathbf{U}_j\}$ to quantify the amount of residual redundancy inherent in each of the other set of CELP parameters (in addition to the LSP parameters).

- **Model A** assumes that the CELP parameters within two different frames are completely independent; it models the intra-frame redundancy using a first-order Markov process. More specifically, it assumes that

$$Pr(\mathbf{U}_j = \mathbf{u}_j | \mathbf{U}_{j-1} = \mathbf{u}_{j-1}, ..., \mathbf{U}_1 = \mathbf{u}_1) = Pr(\mathbf{U}_j = \mathbf{u}_j), \tag{1}$$

and

$$\begin{aligned} Pr(U_{i,j} = u_{i,j} | U_{i-1,j} = u_{i-1,j}, ..., U_{1,j} = u_{1,j}) &= Pr(U_{i,j} = u_{i,j} | U_{i-1,j} = u_{i-1,j}) \quad (2) \\ &\triangleq P_A^{(i)}(u_{i,j} | u_{i-1,j}), \quad (3) \end{aligned}$$

[1]For the LSP's, $l = 10$. For each of the other sets of CELP parameters, $l = 4$.

where $i = 1, 2, \cdots, l$ and $j = 1, 2, \cdots$. Note that for i=1, equation (3) becomes $P_A^{(1)}(u_{1,j})$.

- **Model B** assumes a second-order Markov process: $U_{i,j}$ is independent of all previous parameters conditioned on the immediately preceding parameter $U_{i-1,j}$, and the corresponding parameter in the previous frame $U_{i,j-1}$. This models both the interframe and intra-frame redundancies present in one frame CELP encoded speech. More precisely, we have

$$\begin{aligned} & Pr(U_{i,j} = u_{i,j} | \mathbf{U}_{j-1} = \mathbf{u}_{j-1}, ..., \mathbf{U}_1 = \mathbf{u}_1, U_{1,j} = u_{1,j}, ..., U_{i-1,j} = u_{i-1,j}) \\ = \; & Pr(U_{i,j} = u_{i,j} | U_{i-1,j} = u_{i-1,j}, U_{i,j-1} = u_{i,j-1}) \qquad (4) \\ \triangleq \; & P_B^{(i)}(u_{i,j} | u_{i-1,j}, u_{i,j-1}), \qquad (5) \end{aligned}$$

where $i = 1, 2, \cdots, l$ and $j = 1, 2, \cdots$. Note that for i=1, (5) becomes $P_B^{(1)}(u_{1,j}|u_{1,j-1})$.

The assumption that there exist intra-frame and inter-frame redundancies within the CELP 1016 parameters is based on some of the features of the vocoder. One significant feature is an adaptive smoother, which employs both interpolation of reliable data from neighboring subframes and extrapolation from previous frames. A second feature is the ordered nature of the LSP parameters within one CELP frame ($LSP1 \; < \; LSP2 \; < \; ... \; LSP10$) which suggests intraframe dependency among the LSP's [2].

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

$$H(\mathcal{U}) = \lim_{n \to \infty} H(\mathbf{U}_n | \mathbf{U}_{n-1}, \mathbf{U}_{n-2}, .., \mathbf{U}_1). \qquad (6)$$

$H(\mathcal{U})$ 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 [2, 3]

$$\rho_T \triangleq log_2|\mathcal{U}| \; - \; H(\mathcal{U}), \qquad (7)$$

where $|\mathcal{U}|$ is the size[2] 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:

$$\rho_T = \rho_D + \rho_M, \tag{8}$$

where

$$\rho_D \stackrel{\triangle}{=} log_2|\mathcal{U}| \; - H^*, \tag{9}$$

$$\rho_M \stackrel{\triangle}{=} H^* - H(\mathcal{U}), \tag{10}$$

and $H^* = \sum_{i=1}^{l} H(U_{i,j})$, [2].

A large training sequence (83,826 frames) from the TIMIT speech database [19] 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 set of parameters were compiled to compute its Markov transition probabilities for both Model A and B. These probabilities were used in equations (9) and (10) to calculate ρ_D and ρ_M, respectively. The results for both models are compiled in Table 2 and Table 3, respectively[3]. 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 for the 2^{nd}-order Markov model around 12.5 bits of the 30 high-order bits of the LSP parameters are redundant. If we calculate the total frame redundancy, we find that for Model A among the 78 high-order bits of the CELP parameters, 17 bits (or $\approx 22\%$) of them are redundant. If we add the inter-frame redundancy quantified in Model B, we obtain that 21 bits (or $\approx 27\%$) are redundant.

[2] For the LSP's, $|\mathcal{U}| = 2^{30}$. For each of the other set of CELP parameters, $|\mathcal{U}| = 2^{12}$.

[3] Note that for the pitch delay redundancy, the odd sub-frame pitch delays and the even sub-frame pitch delays are different in nature. Hence, for both models we assume that the pitch delays are independent within a frame. Thus, Model A for the pitch delays consists only of the redundancy due to non-uniformity ρ_D, and Model B consists of the redundancy due to non-uniformity ρ_D and the inter-frame redundancy ρ_M.

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 [22, 23, 4]. We therefore propose to employ unequal error protection (UEP) in order to allow various levels of protection for different parameters. Our UEP system consists of a family of punctured convolutional codes [7], known as rate compatible punctured convolutional (RCPC) codes [14].

Punctured convolution codes were introduced to achieve 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 obtained 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 $\delta \in [1, (n-1)P]$ [16]. For example, using a rate 1/2 mother code and a puncturing period $P = 4$, four different code rates can be attained: $R = 4/5,\ 4/6,\ 4/7$ or $4/8$, where the last rate corresponds to the unpunctured mother code.

The puncturing matrices simply contain 0's, which specify the punctured (or not transmitted) output bits, and 1's, which specify the unpunctured bits. A column of an (n x P) puncturing matrix, $\mathbf{A}(\delta)$, represents the puncturing rule of all the n output streams at a given time (modulo P).

This is best shown by an example. Consider a rate $R = 1/2$ mother code with constraint length 3, and generator matrix $\mathbf{G}(D) = [1 + D^2,\ 1 + D + D^2]$. Let the puncturing period be

$P = 2$, and the puncturing matrix be given by

$$\mathbf{A}(\delta)_{2\text{x}2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{11}$$

This puncturing rule essentially means that we delete every third output bit. The punctured trellis of the above code can be seen in Figure 1. The output bits replaced by the symbol 'X' correspond to the punctured bits. Thus, the resulting code is of rate $R = 2/3$ with an equivalent generator given by

$$\mathbf{G}(D) = \begin{bmatrix} 1+D & 1+D & 1 \\ 0 & D & 1+D \end{bmatrix}. \tag{12}$$

The non-punctured trellis of a rate 2/3 code given by (12) is given in Figure 2. It is easily seen that these two trellises produce the same convolutional code. The only difference lies in the fact that the trellis corresponding to the punctured code is based on that of a rate 1/2 code, which is less complex. Although, the above example is simple, it is evident that many higher rate codes can be attained from a single mother code by using the same low-rate encoder trellis, and thus limiting the complexity of the Viterbi decoding algorithm.

Rate-compatible punctured convolutional (RCPC) codes are a sub-class of punctured codes [14]. 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. Mathematically, it can be understood as follows. Consider a rate $R = 1/n$ mother code with period P, and

$$\mathbf{A}(\delta) = \begin{bmatrix} a_{11}(\delta) & \dots & a_{1P}(\delta) \\ \vdots & \ddots & \vdots \\ a_{n1}(\delta) & \dots & a_{nP}(\delta) \end{bmatrix}, \tag{13}$$

where $\{a_{ij}(\delta) \in \{0,1\} \mid 1 \leq i \leq n, 1 \leq j \leq P\}$, and $1 \leq \delta \leq (n-1)P$. Now, the rate compatibility restriction simply states that:

$$\text{If } a_{ij}(\delta) = 1 \text{ then } a_{ij}(\epsilon) = 1 \text{ for all } \epsilon \geq \delta \geq 1. \tag{14}$$

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 [14].

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. Hence in Figure 1, the places where an 'X' occurs are set to zero in the calculation of the Viterbi decoding metric. Thus, a family of RCPC codes, corresponding to a period P, can be decoded with the same trellis, as long as the different rates (due to different δ's), the corresponding puncturing matrices $\mathbf{A}(\delta)$, and the bits they protect are known at the decoder [16].

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 employ a maximum a-posteriori (MAP) soft-decision decoder that exploits the CELP residual redundancy in combating channel noise. This decoder, which is based on the Viterbi algorithm, chooses the code sequence $\hat{\mathbf{x}}^K = (\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_K)$ that minimizes

$$Pr(\mathbf{y}^K \mid \hat{\mathbf{x}}^K)\ Pr(\hat{\mathbf{x}}^K), \tag{15}$$

where $\mathbf{y}^K = (\mathbf{y}_1, \ldots, \mathbf{y}_K)$ is the received sequence of length K, which is the number of CELP parameters transmitted.

We consider BPSK-modulated AWGN and fully interleaved Rayleigh Fading channels with noise variance $N_0/2$. Thus, the above metric reduces to choosing $\hat{\mathbf{x}}^K$ that minimizes

$$
\begin{aligned}
& \textstyle\sum_{k=1}^{K} \| \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \|^2 \; - \; N_0 \; ln \; Pr(\hat{\mathbf{x}}^K) \\
= \; & \textstyle\sum_{k=1}^{K} \left[\| \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \|^2 \; - \; N_0 \; ln \; Pr(\hat{\mathbf{x}}_k | \hat{\mathbf{x}}_{k-1}, \hat{\mathbf{x}}_{k-2}, ...) \right] \qquad (16) \\
= \; & \textstyle\sum_{k=1}^{K} \left[\| \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \|^2 \; - \; N_0 \; ln \; Pr(\hat{u}_k | \hat{u}_{k-1}, \hat{u}_{k-2}, ...) \right] \quad , \qquad (17)
\end{aligned}
$$

where $\mathbf{a}_K$ is the sequence of Rayleigh fading coefficients which we assume to be available at the decoder. Realize that for the AWGN channel, $\mathbf{a}_k$ is the all-one vector for all k.

In our experiments, $Pr(\hat{\mathbf{u}}^K)$ is calculated using the Markov models of the previous section in conjunction with a large training sequence from the TIMIT database [19]. It is also important to note that the i^{th} quantized CELP parameter in frame j, $u_{i,j}$, is equal to u_k if and only if $k = \phi * j + i$, where ϕ is the total number of transmitted CELP parameters per frame. Note that in the following section two systems will be presented: one that will only encode the LSP parameters ($\phi = 10$) and the other will encode and transmit all the CELP parameters ($\phi = 26$).

In addition, as Section 4 will describe in detail, we will be comparing our UEP system to both an uncoded system and an equal error protection (EEP) system [2] which utilizes a 32-state rate-3/4 convolutional code. This rate EEP system also employs MAP decoding as described by (17). However, since our RCPC UEP system will be using a mother code rate of 1/3, the Viterbi metric must be modified to allow our MAP decoding method to be used[4]. Again, this is due to the fact that the encoder accepts its inputs one bit at a time, while our Markov models are based on 3-bit codewords. This modification simply computes the Viterbi metric during decoding every three trellis steps instead of every one step as is usually the case for a rate 1/3 convolutional encoder. Thus for the RCPC UEP scheme, the decoding metric

[4]The result of this modification is a slight increase in the UEP decoder complexity, which has an equivalent decoding complexity as the EEP scheme [20].

becomes

$$\sum_{k=1}^{K}\sum_{l=1}^{3}\Big[\parallel \mathbf{y}_{k,l}-\mathbf{a}_{k,l}\hat{\mathbf{x}}_{k,l}\parallel^2\Big] \; - \; N_0 \; ln \; Pr(\hat{u}_k|\hat{u}_{k-1},\hat{u}_{k-2},...), \tag{18}$$

where $\mathbf{y}_{k,l}, \mathbf{a}_{k,l}$ and $\hat{\mathbf{x}}_{k,l}$ are the l^{th} bits of the k^{th} received codeword, the k^{th} Rayleigh fading coefficient and the k^{th} estimated codeword, respectively.

As in [2], we use one hard decision and two soft-decision decoding schemes based on the above modified Viterbi metric:

- **ML:** Maximum likelihood[5] decoding which chooses the code sequence $\{\hat{\mathbf{x}}_{k,l}\}$ where $1 \leq l \leq 3$ that minimizes

$$\sum_{k=1}^{K}\sum_{l=1}^{3}\parallel \mathbf{y}_{k,l}-\mathbf{a}_{k,l}\hat{\mathbf{x}}_{k,l}\parallel^2. \tag{19}$$

- **MAP1:** Maximum *a posteriori* decoding exploiting the residual redundancy in Model A, by choosing $\{\hat{\mathbf{x}}_{k,l}\}$ to minimize

$$\sum_{k=1}^{K}\sum_{l=1}^{3}(\parallel \mathbf{y}_{k,l}-\mathbf{a}_{k,l}\hat{\mathbf{x}}_{k,l}\parallel^2) \; - \; N_0 \; ln \; P_A^{([k \bmod \phi])}(u_k|u_{k-1}), \tag{20}$$

  where $[k \bmod \phi]$ refers to the unique integer between 1 and ϕ.

- **MAP2:** Maximum *a posteriori* decoding scheme that exploits the residual redundancy in Model B, by choosing $\{\hat{\mathbf{x}}_{k,l}\}$ to minimize

$$\sum_{k=1}^{K}\sum_{l=1}^{3}(\parallel \mathbf{y}_{k,l}-\mathbf{a}_{k,l}\hat{\mathbf{x}}_{k,l}\parallel^2) \; - \; N_0 \; ln \; P_B^{([k \bmod \phi])}(u_k|u_{k-1},u_{k-\phi}). \tag{21}$$

[5]Note that ML decoding does not exploit any source redundancy; so we could have used the regular Viterbi metric for decoding. However, by using the modified Viterbi metric both our EEP and UEP systems will have the same decoding complexity/delay.

4 Overall System Model

The diagram of the overall system proposed for UEP channel coding of the CELP parameters is shown in Figure 3. The first step is the CELP encoder 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 coding systems: uncoded, equal error protection (EEP) using a 32-state rate 3/4 convolutional code [17], and a 32-state base rate 1/3 RCPC code with period $p = 8$ [16, 20]. We apply our coding systems to two different scenarios.

- Scenario 1 assumes that only the LSP parameters are affected by channel noise, and hence coding is only applied to the 10 LSP parameters of each frame.
- Scenario 2 assumes that all the parameters (excluding the Hamming and synchronization bits) in a CELP frame are affected by noise; thus coding is applied to the entire frame.

The next block in Figure 3 is the BPSK modulation, followed by the channel transmission. In the simulations, two channels are 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, ML or MAP soft-decision decoding is performed based on Models A and B, respectively. The modified Viterbi algorithm of (18) is used to decode the UEP systems, while the metric of (17) is used for the EEP system. The final step is the synthesis of the speech from the decoded parameters.

4.1 Transmission of the LSP Parameters

Four of the ten LSP parameters per frame are 4 bits in length. To be consistent with our previous Markov models (based on the 3 MSB's), only 30 of the 34 LSP bits per CELP frame are convolutionally coded. More specifically, the 4th bits are sent uncoded and hard decision decoded for all transmission schemes.

Through objective and subjective testing, it was found in [22] that the lower LSP's in a CELP 1016 frame are more important to speech reconstruction than the higher ones. The UEP encoder we use is based on a RCPC coded developed in [14], with a rate 1/3, 32-state mother code. Various levels of protection are be applied to the different LSP parameters as shown in Table 4. Note that the last LSP parameter is sent uncoded. However, this parameter was modeled for its residual redundancy. Thus, the MAP soft-decision algorithm can still be applied in the decoding phase. Also, since the LSP parameters exhibit an ordering property, they undergo re-ordering after they are decoded.

4.2 Transmission of all the CELP Parameters

In our EEP and UEP schemes for the transmission of all the CELP parameters, only 78 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 60 bits are sent uncoded and hard decision decoded for all transmission schemes. We assume that the 4 Hamming correction bits, the synchronization bit, and the future expansion bit are sent uncorrupted, since they play no significant role in CELP speech reconstruction.

The same RCPC family of codes are used, and the various coding rates are described in Table 5. They are chosen based on the CELP parameters sensitivity study in [22]. Note that the 3 MSB's of the pitch gain parameters are sent uncoded but can still be MAP decoded since they were modeled for their redundancy.

5 Experimental Results

A large training sequence ($\approx$ 42 minutes) of speech was used from the TIMIT database [19] 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. All the simulations were performed using a practical decoding delay of one frame in length (30 ms). The performance criteria used are:

- The average speech distortion measure, which is an average of seven different speech distortion measures of two different types - cepstral and cosh measures [12]. For a detailed description of each distortion measure refer to [20][Appendix A] and [12]. This distortion measure is averaged over all subframes where those subframes with either zero signal or noise energy are excluded. Note that the minimum possible average speech distortion possible (when the channel is noiseless) is 4.79 dB.

- The symbol error rate, P_s, which is the percentage of parameters in error.

- Subjective listening tests that make pairwise comparisons between the different coding schemes.

The various coding systems used all have different overall rates. To facilitate comparisons of the results for the different systems the following equation was used.

$$\frac{E_b}{N_0} = \frac{1}{R}\frac{E_s}{N_0}, \tag{22}$$

where, E_b refers to the *energy per information bit*, E_s is the *energy per symbol*, and R is the overall code rate.

5.1 Simulation Results for Transmission of LSP Parameters

In [2], the redundancy in the LSP parameters was quantified and exploited through soft-decision MAP decoding. We herein further improve the performance of the system by using our RCPC-based UEP coding scheme with MAP decoding. Note that our UEP scheme has a rate of $\frac{34}{66}$, while the EEP scheme of [2] has a rate of $\frac{34}{44}$. Thus, for comparing the two schemes we use equation (22) and provide the performance for different values of E_b/N_0.

The performance in terms of average speech distortion and symbol error rate of the EEP and UEP systems with ML/MAP decoding over AWGN and Rayleigh fading channels are shown in Figures 4 to 7. The results for an uncoded system are also presented for reference. It can be clearly remarked from all the figures that the UEP-MAP2 scheme – which exploits both intra-frame and inter-frame LSP redundancies – substantially outperforms all the other schemes; it also offers a very graceful degradation as the channel conditions deteriorate. In particular, at an average speech distortion of 6.0 dB over the AWGN channel, UEP-MAP2 achieves a gain of 1.05 dB over EEP-MAP2. Over the Rayleigh channel, the gain is 1.88 dB. Furthermore, for the same speech distortion, EEP-MAP2 performs better than EEP-ML by 3.27 dB over the AWGN channel and by 5.64 dB over the Rayleigh channel. This results in an overall gain for UEP-MAP2 versus EEP-ML of 4.32 dB over the AWGN channel and of 7.52 dB over the Rayleigh channel.

Additional results using the spectral distortion criterion [2] are obtained in [20], and similar improvements are observed. Furthermore, the performance of other EEP and UEP schemes with various code rates is evaluated in [20]. For these systems, at an average speech distortion of 6.0 dB, the gains due to UEP-MAP2 over EEP-MAP2 range from 0.3 dB to 1.33 dB over the AWGN channel, and from 0.5 dB to 2.37 dB over the Rayleigh channel. The gains of UEP-MAP2 over UEP-ML, at the same speech distortion, vary from 2.32 dB to 3.07 dB over the AWGN channel, and from 2.61 dB to 4.37 dB over the Rayleigh channel.

5.2 Simulation Results for Transmission of all the CELP Parameters

Using the residual redundancy present in all the CELP parameters, soft-decision MAP decoding is now applied to the entire frame of CELP parameters. In addition, the gains achieved when unequally protecting different CELP parameters using RCPC codes is examined.

As described in Section 4.2, the UEP scheme for the overall system produces rate of $\frac{138}{252}$, while the EEP scheme of [2] has a rate of $\frac{138}{162}$. Once again, equation (22) is used to compare the performance of the two schemes. The performance of the two schemes with ML/MAP decoding, as well as that of the uncoded scheme are shown in Figures 8 and 9 for various values of E_b/N_0. As in the case of the LSP transmission, it can be clearly observed from the figures that the UEP-MAP2 scheme provides the best performance. At an average speech distortion of

6.0 dB, the gains for UEP-MAP2 versus EEP-MAP2 are 1.42 dB and 2.55 dB over the AWGN and Rayleigh channels, respectively. This is 0.37 dB and 0.93 dB larger than the results for the same systems protecting only the LSP's. Furthermore, the gains for EEP-MAP2 versus EEP-ML are 1.82 dB and 2.47 dB over the AWGN and Rayleigh channels, respectively. This results in an overall gain for UEP-MAP2 versus EEP-ML of 3.24 dB over the AWGN channel and of 5.02 dB over the Rayleigh channel.

Finally, it can be observed that for low to medium values of E_b/N_0, the UEP-MAP2 scheme provide significant gains over the EEP-MAP2 scheme. However, at high values of E_b/N_0, the EEP-MAP2 scheme provides a better performance (due to the asymptotic coding gain obtained by channel coding all the parameters). This suggests that when the system is operating at low to medium channel signal-to-noise ratios (SNR), it is recommended to use the UEP-MAP2 scheme; then as the channel conditions improve, the system can switch to the EEP-MAP2 scheme. This results in an overall *adaptive and robust* system which can estimates the channel error conditions and select accordingly an appropriate error protection scheme (UEP at low SNR's and EEP at high SNR's).

5.3 Listening Tests

We performed listening tests that made pairwise comparisons between the EEP and UEP schemes for the transmission of all the CELP parameters over the Rayleigh fading channel using MAP2 decoding. The tests were obtained for 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. Then a pair of different system outputs at the same E_b/N_0 was played, and the listener was asked to choose which one sounded better, without being told which system corresponded to which segment. If they failed to notice a significant difference, they were given the option to choose *neither*. Two pairwise comparisons were made at each E_b/N_0 by each listener. The results of the tests are shown in Table 6. We draw the following observations.

- At low SNR, $E_b/N_0 = -2$ dB, the UEP scheme performs significantly better than the EEP scheme, with 96% selecting UEP over EEP for the first speech segment and 84% for the second. No listeners chose EEP as the better system at this low SNR.

- When the SNR was increased to $E_b/N_0 = 1$ dB, the results showed that the UEP system once again outperformed the EEP scheme but not as significantly as at the lower SNR. This time 70% and 90% of the listeners for speech segments 3 and 4, respectively, selected the UEP system over the EEP system. However for speech segment 3, 14% of the listeners preferred the EEP scheme over the UEP scheme.

All the systems use MAP2 decoding; thus these tests evaluate the effect of unequal error protection on the quality of speech reconstruction. The results showed that the UEP system clearly performed better than the EEP system of [2]. For a demonstration of the listening 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. Two different Markov models were proposed to generate the CELP parameters and to quantify the amount of residual redundancy they exhibit both within a frame and between frames. 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) MAP soft-decision detection that utilizes the CELP residual redundancy in combating channel noise. The system was first applied to the transmission of the LSP parameters, and then applied for the transmission of all the CELP parameters. Experimental results also showed 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.

This study could also be extended to other similar low-bit rate vocoders [8]. It would be interesting to see how the protection schemes explored in this paper would apply to these newer vocoders. Finally, we should point out that other methods of unequal error protection exist, such as the application of different levels of protection through transmission energy allocation [1, 11]. In this method different energy levels are allocated relative to the importance of the specific parameter in the frame. This method can be applied to our system in conjunction with our proposed MAP RCPC-based schemes.

References

[1] F. Alajaji, S. Al-Semari and P. Burlina, "An Unequal Error Protection Trellis Coding Scheme for Still Image Communication," *Proc. Int. Symp. on Inform. Theory*, Ulm, Germany, June 1997.

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

[3] F. Alajaji, N. Phamdo, N. Farvardin and T. Fuja, "Detection of Binary Markov Sources Over Channels with Additive Markov Noise," *IEEE Trans. Inform. Theory*, Vol. 42, No. 1, pp. 230-239, January 1996.

[4] E. Bracha, N. Farvardin and Y. Yesha, "Channel Coding for CELP Encoded Speech in Mobile Radio Applications," Digital Signal Processing Group, Advanced Systems Division, NIST Report, August 1994.

[5] P. Burlina and F. Alajaji, "An Error Resilient Scheme for Image Transmission Over Noisy Channels with Memory," *IEEE Trans. on Image Processing*, Vol. 7, pp. 593-600, April 1998.

[6] J. P. Campbell, T. E. Tremain and V. C. Welch, "The DoD 4.8 kbps Standard (Proposed Federal Standard 1016)," *in Advances in Speech Coding,* Kluwer Academic Publishers, Boston, pp. 121-133, 1991.

[7] 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. Inform. Theory*, vol. 25, pp. 97-100, January 1979.

[8] R. V. Cox, "Three New Speech Coders from the ITU Cover a Range of Applications," *IEEE Communication Magazine*, pp. 40-47, September 1997.

[9] N. Farvardin, "Speech Coding over Noisy Channels," *Institute for Systems Research*, Tech. Rep. TR-94-5, University of Maryland, College Park, MD, 1994.

[10] T. Fazel and T. Fuja, "Channel-Encoded Transmission of MELP-Compressed Speech," *Proc. Int. Symp. Inform. Theory*, MIT, Cambridge, MA, August 1998.

[11] S. Gadkari and K. Rose, "Transmission Energy Allocation with Low Peak-to-Average Ratio," *IEEE Communications Letters*, Vol. 1, pp. 166-168, November 1997.

[12] A. H. Gray and J. D. Markel, "Distance Measures for Speech Processing," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-2 4, no. 8, Oct. 1976.

[13] J. Hagenauer, "Source Controlled Channel Decoding, " *IEEE Trans. on Commun.*, Vol. 43, No. 9, pp. 2449-2457, September 1995.

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

[15] P. Hedelin, P. Knagenhjelm and M. Skoglund, "Vector Quantization for Speech Transmission," Chapter 9, in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, ed., Elsevier Science, 1995.

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

[17] S. Lin and D.J. Costello, Jr., *Error Control Coding*, Prentice Hall, 1983.

[18] 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.

[19] National Institute of Standards and Technology (NIST), "The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM," NIST, October 1990.

[20] 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, Nov. 1997. Available at: *http://markov.mast.queensu.ca/publications.html.*

[21] N. Phamdo, N. Farvardin and T. Moriya, "Combined Source-Channel Coding of LSP Parameters Using Multi-Stage Vector Quantization," in *Speech and Audio Coding for Wireless and Network Applications*, B. S. Atal *et al.* ed., Kluwer, Boston, pp. 181-190, 1993.

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

[23] D. E. Ray and D. J. Rahikka, "Reed-Solomon Coding for CELP EDAC in Land Mobile Radio," *Proc. ICASSP*, Adelaide, Australia, April 1994.

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

[25] M. Skoglund, "Fixed and Adaptive VQ Decoding in Robust LPC Quantization," *Proc. IEEE Work. on Speech Coding and Telecommun.,* pp. 71-72, Annapolis, September 1995.

[26] H. Xiao and B. Vucetic, "Soft Input Source Decoding in Low-Bit-Rate Speech Transmission," *Proc. Int. Symp. on Inform. Theory*, Ulm, Germany, June 1997.

	Spectrum	Adaptive	Stochastic
Update Every	30 ms (240 samples)	$30/4 = 7.5$ ms (60 samples)	$30/4 = 7.5$ ms (60 samples)
Bits per Frame	34 LSP bits [3444433333]	index: $8+6+8+6$ gain: 5×4	index: 9×4 gain: 5×4
Note: The remaining 6 bits are used as follows: 1 bit per frame for synchronization, 4 bits per frame for forward error correction and 1 bit per frame for future expansions.			

Table 1: Bit Allocation in a FS CELP 1016 Encoded Frame of Speech.

CELP Parameter	Redundancy		
	ρ_D	ρ_M	ρ_T
LSP	5.2747	4.5927	9.8674
Codebook Gain	4.0478	1.024	5.0718
Pitch Gain	0.1832	1.1335	1.3167
Pitch Delay	0.7064	0.0000	0.7064
Codebook Index	0.0323	0.0181	0.0504
Total Frame	10.2444	6.7683	17.0127

Table 2: CELP 1016 Redundancy (in Bits/Frame) using Model A.

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 3: CELP 1016 Redundancy (in Bits/Frame) using Model B.

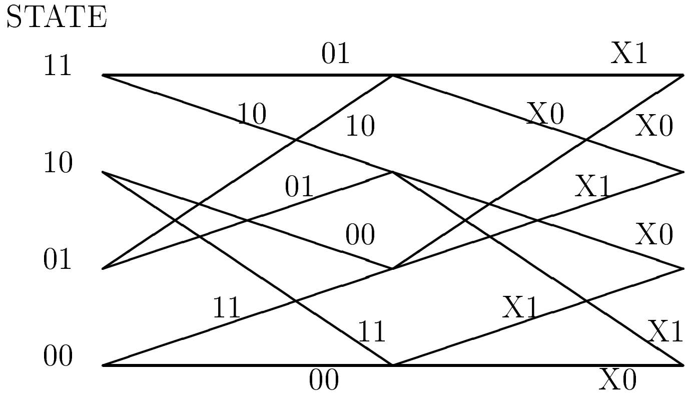

Figure 1: Trellis for Rate $R = 1/2$, Constraint Length 3, Period $P = 2$ Punctured Convolutional Code.

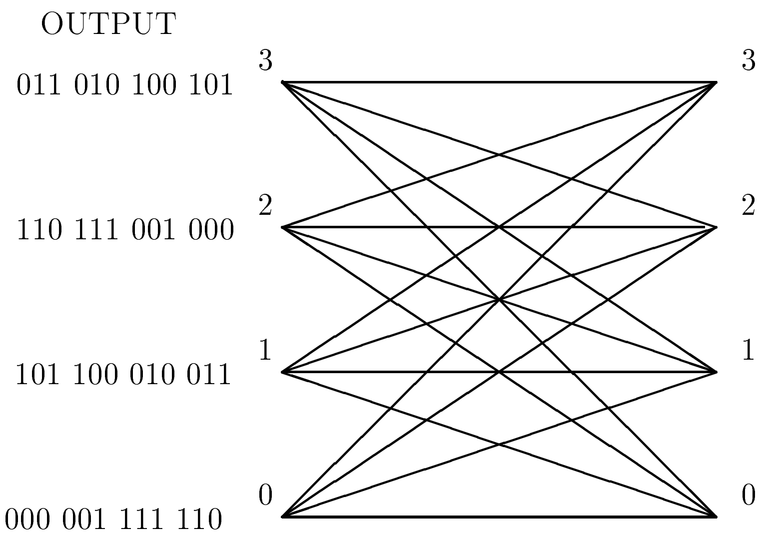

Figure 2: Trellis for Rate $R = 2/3$, Constraint Length 3, Non-punctured Convolutional Code.

LSP	Code Rate
1	8/20
2-5	8/18
6-9	8/16
10	Uncoded

Table 4: LSP Unequal Error Protection Scheme Using Mother Rate-1/3 Family of RCPC Codes.

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 5: Overall CELP UEP Scheme Using Mother Rate-1/3 Family of RCPC Codes.

$E_b/N_0 = -2$ dB	Speech Segment 1	UEP: 96 % EEP: 0% Neither: 4%
$E_b/N_0 = -2$ dB	Speech Segment 2	UEP: 84 % EEP: 0% Neither: 16%
$E_b/N_0 = 1$ dB	Speech Segment 3	UEP: 70% EEP: 14% Neither: 16%
$E_b/N_0 = 1$ dB	Speech Segment 4	UEP: 90% EEP: 0% Neither: 10%

Table 6: Listening Test Results for UEP vs EEP over the Rayleigh Fading Channel using MAP2 Decoding.

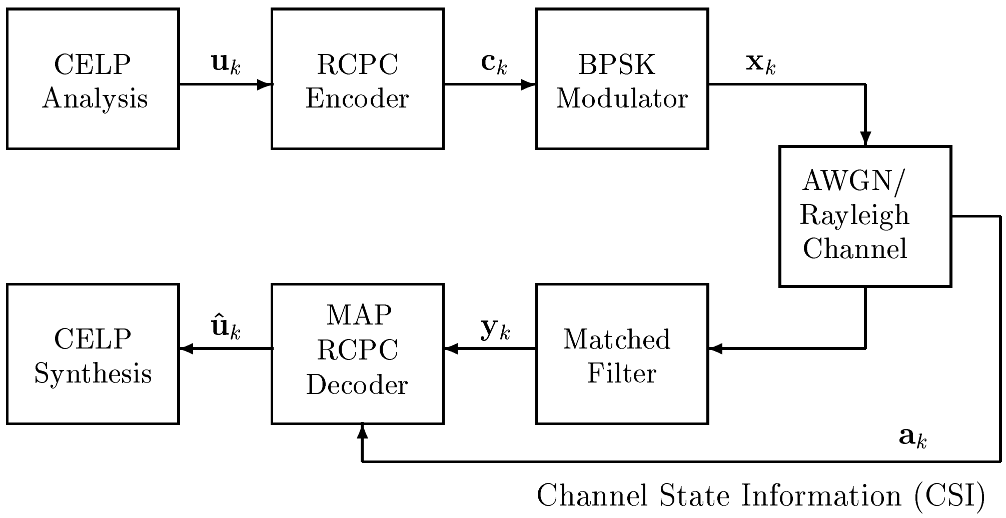

Figure 3: Block Diagram of the Overall System with MAP RCPC Decoder.

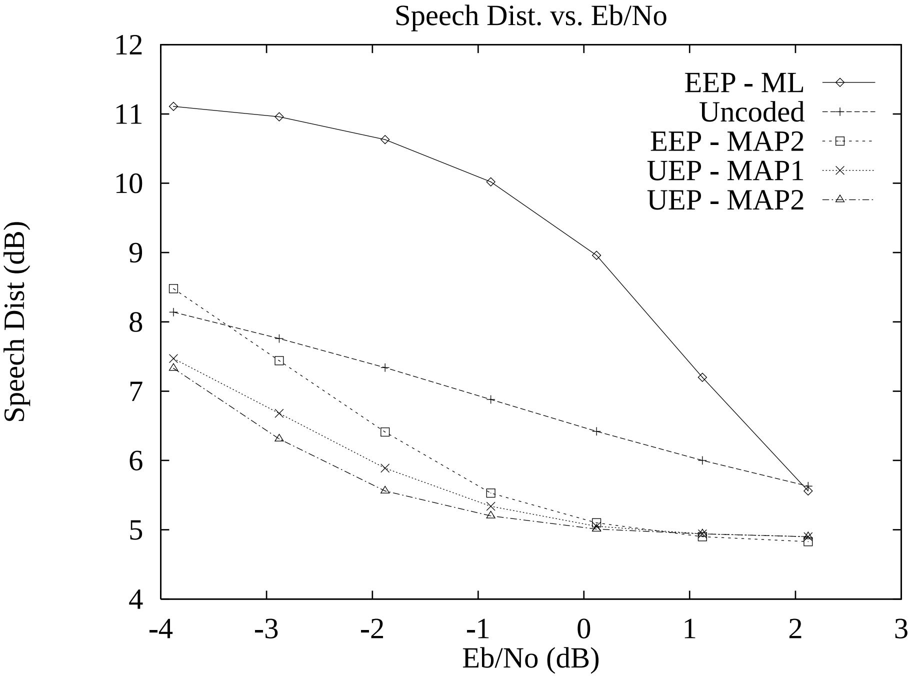

Figure 4: Average Speech Distortion for Different Coding Schemes of the LSP Parameters over the AWGN Channel.

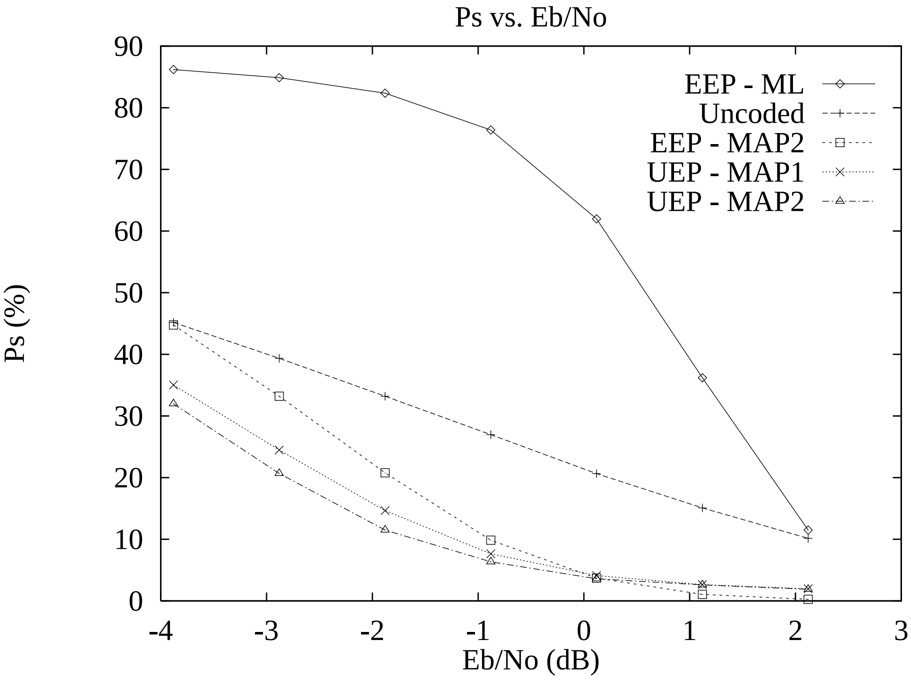

Figure 5: Symbol Error Rates for Different Coding Schemes of the LSP Parameters over the AWGN Channel.

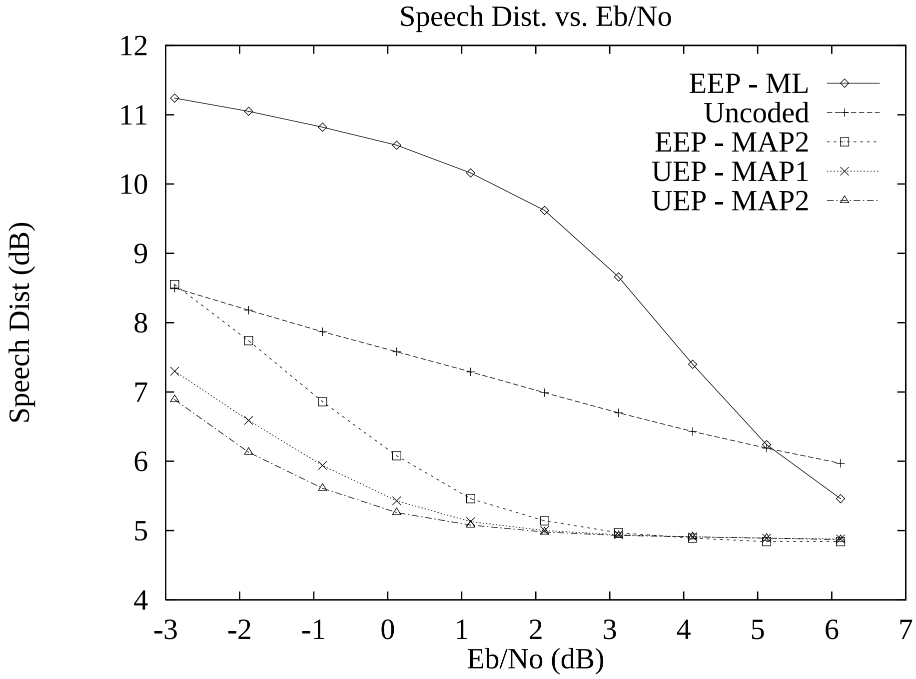

Figure 6: Average Speech Distortion for Different Coding Schemes of the LSP Parameters over the Rayleigh Fading Channel.

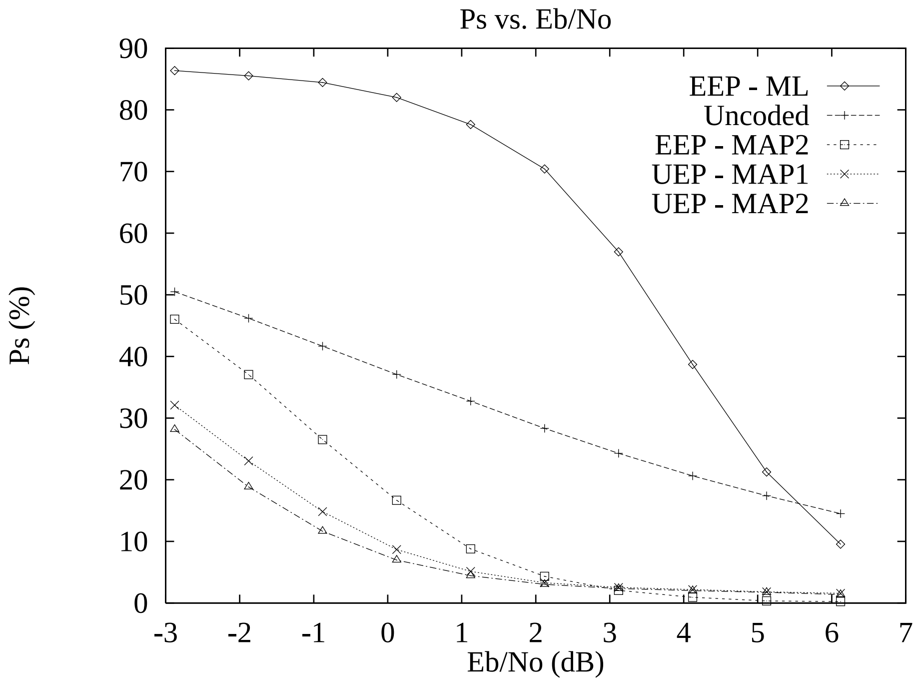

Figure 7: Symbol Error Rates for Different Coding Schemes of the LSP Parameters over the Rayleigh Fading Channel.

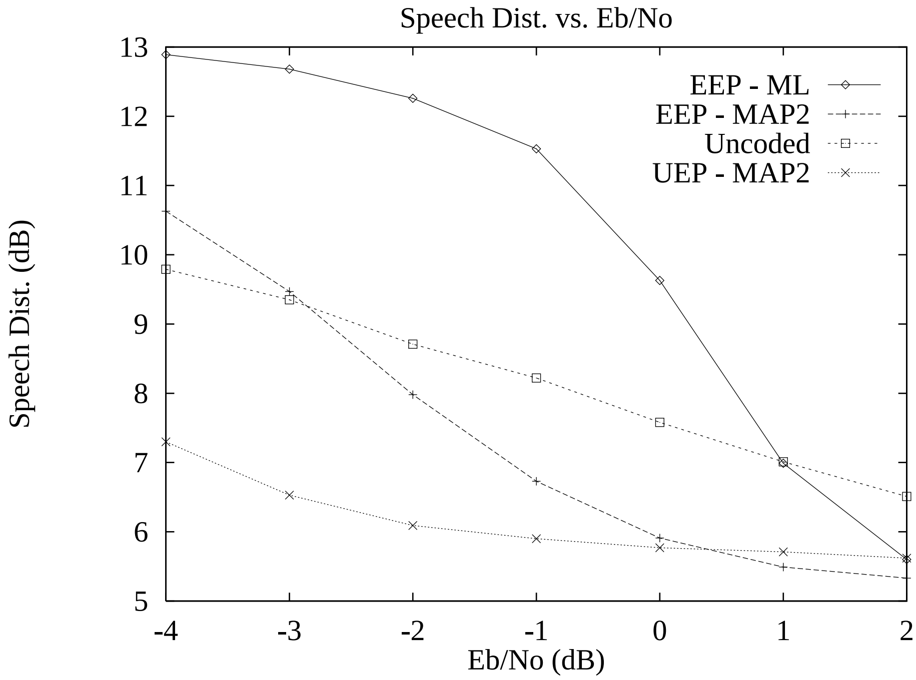

Figure 8: Average Speech Distortion for Different Coding Schemes of all the CELP Parameters over the AWGN Channel.

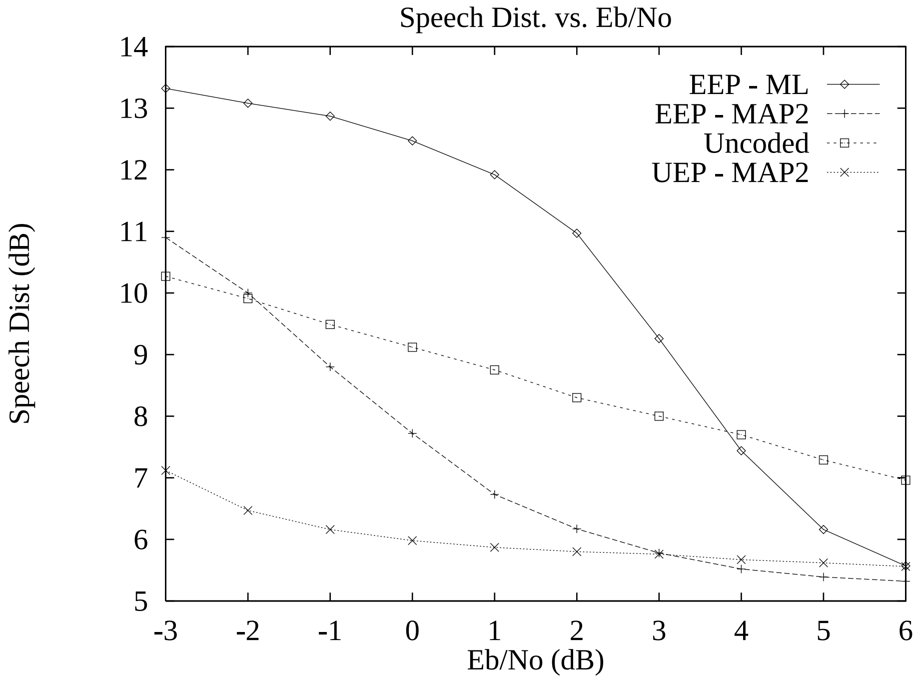

Figure 9: Average Speech Distortion for Different Coding Schemes of all the CELP Parameters over the Rayleigh Fading Channel.