JOINT SOURCE-CHANNEL CODING AND UNEQUAL ERROR PROTECTION OF CELP 1016 ENCODED SPEECH OVER VERY NOISY CHANNELS

by

Ali Nazer

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements
for the degree of Master of Science (Eng.)

Queen's University
Kingston, Ontario, Canada
December 1997

Abstract

The need for a reliable means of wireless communication has become essential in today's society. However, with each new technological advancement come many engineering hurdles. In the wireless environment, constraints on the available bandwidth and channel noise/distortion are two such problems.

In this thesis, we consider the problem of the reliable transmission of speech over very noisy communication channels. More specifically, we investigate joint source-channel coding methods in conjunction with unequal error protection (UEP) for the transmission of Federal Standard CELP 1016 encoded speech. They consist of source-optimized channel coding schemes, which take the statistical characteristics of the source output into consideration during the channel decoding process.

We begin by quantifying the amount of residual redundancy inherent in the CELP 1016 line spectral pair (LSP) parameters. First and second-order Markov chains are introduced to model the LSP correlation within a speech frame and between frames, respectively. A large training sequence from the TIMIT speech database is used in conjunction with these models to quantify the redundancy found within the LSP's. The redundancy is then exploited through a convolutional coding scheme employing *maximum a-posteriori* (MAP) soft decision decoding. This results in a joint source-channel coding system that incorporates the source characteristics in the channel decoder. We then apply unequal error protection to the LSP parameters by means of a rate-compatible punctured convolutional (RCPC) code family; thereby, providing

more protection to the important parameters. Simulations for the communication of the LSP's over the additive white Gaussian noise (AWGN) and the fully interleaved Rayleigh fading channels using binary phase-shift keying (BPSK) demonstrate that the use of UEP and MAP channel decoding are beneficial, particularly during severe channel conditions.

We next extend our results for the case where all the CELP parameters are subject to the effects of channel noise and signal fading. Similar Markov-based models are implemented to quantify the overall residual redundancy inherent in the CELP 1016 parameters. We similarly apply the RCPC-based UEP joint source-channel coding scheme for the transmission of the entire set of CELP parameters. Experimental results over BPSK-modulated AWGN and independent Rayleigh fading channels indicate substantial subjective and objective improvements over traditional equal error protection *maximum likelihood* schemes. More specifically, coding gains up to 7 dB in E_b/N_0 are achieved over systems that do not employ UEP and joint source-channel coding.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Fady Alajaji, for his guidance. Every large project comes with its share of obstacles, and with Dr. Alajaji's dedication and support, I was able to overcome each challenge as it arose.

Next, I would like to thank my peers in the Communications group of Mathematics and Engineering, especially Julian Cheng and Nick Whalen. They were always willing to assist me in my many problems, and without their support this work would not have been possible.

I wish to also thank Dr. Bob Erdahl for giving me the opportunity to join the Mathematics and Engineering program, from which I have gained so much.

Finally, I would also like to acknowledge Queen's University School of Graduate Studies and the National Science and Engineering Research Council for their financial support over the past two years.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The technology to communicate over long distances was first introduced during the nineteenth-century by innovative scientists such as Morse and Bell. Since then, the field of telecommunications has proliferated to encompass not only telegraph or voice signals, but also computer data, images and even video. The communication medium has also changed. Copper lines are no longer the stand-alone means of transmission; we now have the sophistication to transmit information over wireless channels (such as satellite and cellular channels) and optical fiber cables. With these many advancements have come numerous growing pains, which have kept the role of the communications engineer both challenging and dynamic.

The general point-to-point telecommunication system can be modeled as shown in Figure 1.1. The role of the *source encoder* is essentially to transform the input signal into a more compact form. In today's digital environment, this entails changing the input format, such as video, into a binary format. The ideal source encoder compresses the input signal by removing its redundant information; for this reason source encoding is also referred to as source compression. The *channel encoder* essentially performs the opposite operation; it 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. The *modulator* then prepares the signal for transmission, by converting the binary information into signals or pulses, more suitable for transmission. Once the signal has been transmitted over the channel, the opposite of the above functions are performed to produce an estimate of the original signal.

The communications engineer has many constraints to take into account when designing a system for the reliable transmission of information signals. Complexity of the system, transmission delay and available bandwidth are just a few of the limitations placed on any communication scheme.

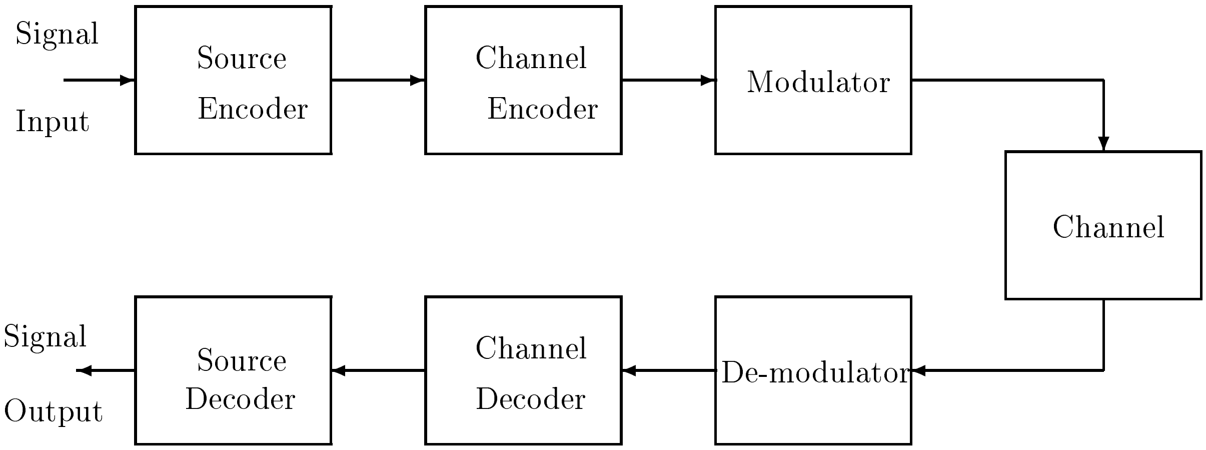

Figure 1.1: Block Diagram for a Telecommunications System

In the above description of a telecommunication system, the source and channel coding were treated as separate entities; this approach is known as *tandem source-channel coding*. This is justified by Shannon's Separation Principle [32] 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.

With the expanding need for reliable wireless communication, there are limitations placed on both the complexity and delay of the system. For instance, the wireless channel is bandlimited. Also, wireless channels are much worse than traditional copper lines, since they exhibit memory and signal fading. With these constraints in mind, researchers have begun to *jointly* design source and channel coding operations. These *joint source-channel coding* schemes have shown better performance than traditional schemes under these limitations.

1.1 Literature Review

In tandem source-channel coding there is no diffusion of knowledge between those implementing the source coder and those involved with the channel coder. In other words, both systems are designed separately, with no knowledge of the other system's specifications. A coding system, in which information about the source code is integrated into the design of the channel code (or vice-versa), can be thought of as a joint source-channel coding method.

At the simplest level, this may consist of identifying which source encoder output bits (or vectors) are important for the proper reconstruction of the signal. This results in an *unequal error protection* (UEP) scheme at the channel encoder, which provides a higher level of protection to the important bits.

Hagenauer developed a type of convolutional code system that can provide different levels of protection through different convolutional code rates. These types of codes are easy to implement and are known as rate-compatible punctured convolutional (RCPC) codes [14]. In [16], Hagenauer applies these RCPC codes to digital mobile radio with a sub-band speech coder. He attains good coding gains using his UEP-RCPC scheme over the correlated Rayleigh fading channel with differentially coherent four-phase shift keying. He also applies RCPC error protection to image transmission problems in [33]. Both papers indicate (objectively and subjectively) that at lower signal-to-noise ratios, unequal error protection schemes significantly outperform equal error protection methods.

David Ray also used unequal error protection in the channel coding of Federal Standard CELP 1016 [25] encoded speech. This was done by applying different types of Reed-Solomon codes for different blocks of CELP 1016 parameters. His results about the performance comparison of unequal error protection versus equal error protection were not conclusive, since no subjective listening tests were performed.

In the case of images, some bits can be left completely unprotected since their importance in image reconstruction is minimal. In [23] and [24], Modestino and Daut employ unequal error protection in the above manner.

At a higher level, joint source-channel coding can involve much more detailed knowledge about the source or channel in question. One such approach is to use the source characteristics in developing the channel code. In the paper by Kroll and Phamdo [18], this source-optimized channel coding method is employed and shown to outperform traditional channel coding designs.

Similarly, the source characteristics can be exploited in the channel decoding phase. Alajaji *et. al.* [2] exploited the residual redundancy present in binary Markov sources. In this situation no channel coding was used, instead the redundancy in the source is used to protect against channel errors. In a later work by the same group, both channel codes and source redundancy are used to combat errors [1]. In [1], the redundancy present in the line spectral pair (LSP) parameters of CELP 1016 encoded speech, is exploited in the channel decoder. Both Reed-Solomon and convolutional codes are employed with BPSK modulation over the AWGN and fully interleaved Rayleigh fading channels. It is found that the use of the LSP characteristics significantly improves the quality of speech, both subjectively and objectively, particularly during very bad channel conditions.

Hagenauer has applied a similar source-optimized channel decoding system to image transmission [33]. He employs a system with a non-adaptive discrete cosine transform source encoder and a convolutional coding scheme that employs a soft-decision Viterbi algorithm that exploits the compressed image characteristics. Substantial gains are observed especially in very noisy environments.

In this thesis, the work of Alajaji *et. al.* in [1] is extended to encompass all the CELP 1016 parameters. More specifically, source-optimized channel decoding of convolutional codes via the maximum a-posteriori (MAP) Viterbi metric is employed for the CELP parameters. Also, unequal error protection through RCPC codes is used in conjunction with MAP decoding for CELP 1016 encoded speech.

1.2 Thesis Outline

The thesis is organized into six chapters including the introduction. In Chapter 2, we briefly discuss the CELP 1016 speech coding standard. We also investigate the different sensitivities of the CELP parameters to channel noise. In Chapter 3, the redundancy in the LSP parameters is modeled through both first order and second-order Markov models. This residual redundancy is then exploited through convolutional codes using the Viterbi MAP metric. In Chapter 4, unequal error protection through RCPC codes is used in conjunction with source-optimized channel decoding on the LSP parameters. In Chapter 5, the residual redundancy in the remaining CELP parameters are quantified, and a UEP-MAP coding scheme is applied to all the CELP 1016 parameters. Chapter 6 summarizes the results and discusses possibilities for future work.

Chapter 2

Code Excited Linear Prediction (CELP)

Code Excited Linear Prediction (CELP) vocoders were developed jointly by AT&T Bell Laboratories and the U.S Department of Defense in 1988 [25]. The Federal Standard 1016 CELP was adopted in order to meet the vocoder requirements of digital Land Mobile Radio Federal Standard 1024 [30]. CELP coding has many other possible applications in both public safety standards and micro-cellular personal communications systems.

2.1 Overview of CELP 1016

The general CELP model consists of three main components: the all-pole linear prediction (LP) filter, the adaptive codebook, and the stochastic codebook. The LP filter is used to model the human vocal tract, while the adaptive and stochastic codebooks are used to generate the *voiced* and *unvoiced* excitations, respectively, as seen in the model for human speech production (refer to Figure 2.1). These are the two types of excitation present in human speech. *Voiced* excitation is produced when quasi-periodic vibrations of the vocal folds interrupt the passage of air from the lungs (e.g., the sound of the letter **i** in s**i**x). On the other hand, *unvoiced* is a turbulent, noise-like excitation caused by a constriction of airflow at some point along the vocal tract (e.g., the sound of the letter **s** in **s**ix) [8].

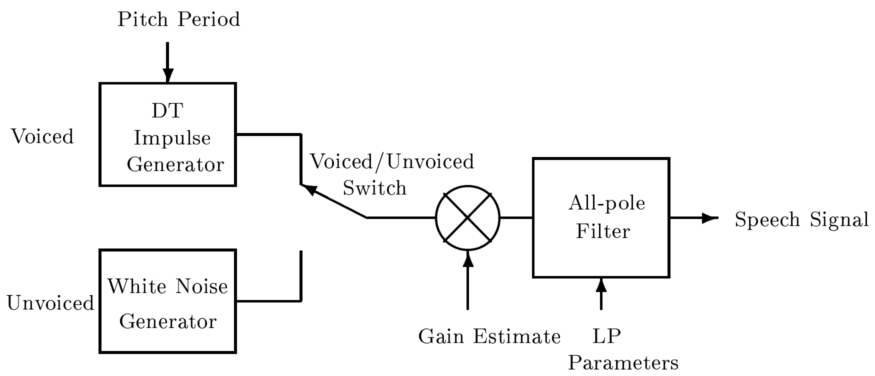

Figure 2.1: Linear predictive model for speech production.

In addition, the model has two phases – analysis and synthesis. The analysis (or coding) phase consists of using the digitized input speech signal to optimally estimate the model parameters (the exact definition of optimal will be discussed below). The synthesis (or decoding) phase is the exact opposite, in which the transmitted parameters are used to re-construct the speech signal at the output; in addition, a postfilter is employed during synthesis to enhance speech quality.

It is important to note that CELP analysis is actually an analysis-by-synthesis method, in which the following functions are performed: 1) estimation of the signal's short-term spectrum through the LP filter co-efficients*; 2) search for an adaptive codebook vector to model the long-term signal periodicity; and 3) vector quantization of the residual from the signal spectrum and adaptive codebook vector using a randomly-generated stochastic codebook. In CELP 1016, both searches are performed to minimize the time-varying, squared prediction error of the perceptually

*In CELP 1016 line spectral pairs (LSP's) are used as an alternate representation of the LP filter co-efficients.

weighted† error signal (the optimal estimation referred to above). Synthesis, simply, performs the above steps in reverse order.

Federal Standard 1016 CELP is a frame-oriented technique that samples the input signal at 8kHZ and breaks the corresponding samples into blocks or vectors, which are then processed as one unit. Each frame (i.e., block), containing 240 samples, is 30ms in duration and produces 144 bits representing the CELP parameters (See Table 2.1). Each frame is further divided into four 7.5ms sub-frames. CELP 1016 has an overall output rate of 4800 bits per second.

	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(-1,2): 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 to provide future expansion(s) of the coder.			

Table 2.1: Bit allocation in a FS CELP 1016 encoded frame of speech.

2.2 Short-Term Linear Predication

The term "short-term linear prediction" is derived from the fact that the all-pole filter gives rise to a system of linear equations, which represent the signal's short-term spectrum, that are then solved to derive the LP coefficients. In CELP 1016,

†Perceptual weighting improves the subjective quality of speech by exploiting the masking properties of human hearing [25].

the LP co-efficients are not used, instead an alternative representation via the *line spectral pairs* (LSP's) is employed. The LSP's refer to pairs of poles in the LP filter. By definition, these poles are located on the unit circle in the z-domain, and are equivalent to undamped sinusoids in the time domain. The fact that these sinusoids have line spectra gives rise to their name – line spectral pairs. It is important to note that since the LSP's lie on the unit circle they have unity magnitude, and thus can be explicitly defined in terms of frequency [8].

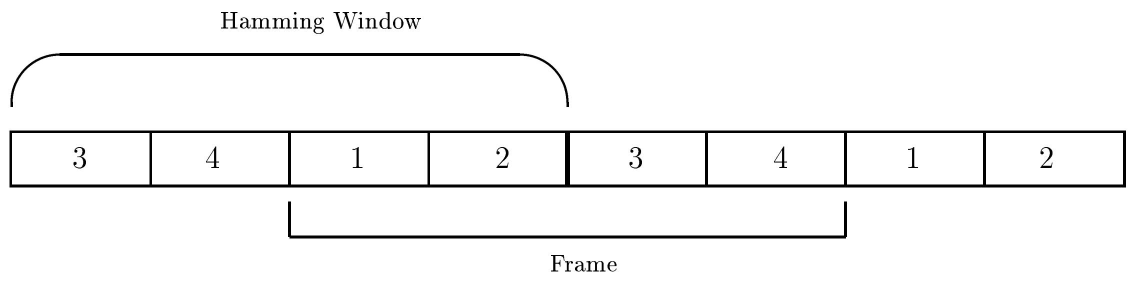

Figure 2.2: Diagram of a CELP 1016 frame.

The LSP's, which are derived from the LP filter coefficients, define the LP filter, and are used to model the signal's short-term spectrum. More specifically, CELP 1016 uses a 10^{th} order, minimum phase, all-pole filter to model the human vocal tract. Open loop, linear prediction through 10^{th} order autocorrelation, in conjunction with a 30ms Hamming window, is used to produce the 10 LSP's per frame from the input speech signal. The Hamming window, which has the same duration as a frame, is centered on the transition between frames. Hence, LP analysis is done only once per frame and is linearly interpolated for each sub-frame using the past and future spectra as in Table 2.2.

It is important to note that the internal delay of the filter is in large part due to the calculation of the LP filter coefficients. Thus, due to the position of the Hamming

Sub-Frame	Past Spectrum	Future Spectrum
1	7/8	1/8
2	5/8	3/8
3	3/8	5/8
4	1/8	7/8

Table 2.2: Interpolating Weights for Sub-Frame LSP's.

window for the future spectrum, there is a 15ms delay in the calculation of the LSP's.

Another important feature of the LSP's is that they are monotonic – higher numbered LSP's must not contain lower frequencies than the lower numbered LSP's in that frame. This can be seen through the derivation of the LSP parameters, and will be discussed in furthur detail in the upcoming section.

In addition, CELP 1016 uses a 15 Hz Bandwidth expansion but no pre-emphasis‡. The bandwidth expansion replaces the LP filter parameters, a_i, by $\gamma^i a_i$, where $\gamma^i = 0.994$ for a 15 Hz expansion. The bandwidth expansion shifts the poles radially inward, thus, improving the speech quality as well as being beneficial in both quantization and search methods used for converting LP coefficients directly into quantized LSP's.

The LSP's undergo independent, non-uniform, scalar quantization as specified by Federal Standard CELP 1016 [25]. This produces 34 bits per frame corresponding to the 10 LSP's, where the first LSP and the last five LSP's are 3-bits long, and the remaining LSP's are 4-bits long.

‡Pre-emphasis filtering increases the relative energy of the high frequency components [8].

2.3 Codebook Excitation

As previously mentioned, both codebook searches are done by a closed loop search using a modified squared prediction error criterion of a time varying, perceptually weighted error signal. In this case, the perceptual weighting filter is similar to a bandwidth expansion where $\gamma^i = 0.8$, corresponding to an expansion of 568 Hz. The overall objective of these searches is to find the best match between the digitized input speech and the CELP synthesized output speech [26].

The adaptive codebook, which is responsible for the quasi-periodic "voiced" excitation, consists of 147 elements. This codebook is simply a shifting register that stores the previous samples from approximately the last 2.5 subframe intervals. Thus, the adaptive codebook excitation vector is a specifed group of 60 previous input samples, which produce the periodic effects needed for "voiced" excitation.

In odd subframes, the adaptive codebook elements to be used as the excitation vector are specified by an 8-bit adaptive codebook index (or *pitch delay*). Thus, there are 256 possible codewords: 128 integer delays, and 128 non-integer delays ranging from 20 to 147. The non-integer delays are realized by 40-point interpolation. To reduce complexity in the even subframes, the pitch delay is 6-bit delta coded relative to the previous 8-bit delay.

In addition there is also a 5-bit adaptive codebook gain (or *pitch gain*), sent each sub-frame, which ranges between -1 and $+2$, and is coded by absolute, non-uniform, scalar quantization according to the standards of CELP 1016 [25].

The stochastic codebook is responsible for the "unvoiced" excitation, which can be modelled by zero mean, unit variance, uncorrelated white noise. Thus, the stochastic codebook contains zero mean, unit variance, ternary (i.e., -1,0, or $+1$) white Gaussian sequences. There are 1082 ternary elements in this codebook, which give rise to 512 overlapped codes of 60 elements (one subframe interval) each. Thus, the stochastic excitation vector (or *codebook index*) is 9 bits in length, and is sent in each sub-frame [26].

As for the adaptive codebook, there is a corresponding stochastic codebook gain (or *codebook gain*), which is 5 bits in length and ranges from -1330 to +1330. It is quantized in a similar way as the pitch gain, and is also sent in each sub-frame.

2.4 Other Features of CELP 1016

There are other important elements in the CELP 1016 vocoder design. Of primary importance are the error protection schemes that are designed to reduce undetected decoding errors, which cause perceptually annoying synthesis errors. These schemes include a Hamming forward error correcting code and an adaptive smoother.

The $(15, 11)$ Hamming code is capable of detecting and correcting one error, and protecting up to 11 bits per frame. In the CELP 1016 vocoder only 10 bits per frame are protected and 1 more bit can be protected in future standards. These protected bits include the 3 most significant bits of the odd sub-frame pitch delays, and the most significant bits of all the pitch gains.

The smoothing technique employs interpolation of reliable parameters from neighboring sub-frames and extrapolation of reliable parameters from previous subframes. The only parameters that cannot undergo adaptive smoothing are the stochastic codebook indices, since they are not correlated or ordered. The highly ordered nature of the monotonic LSP's does allow for linear interpolation [30], a property that gives rise to much of the redundancy present in the LSP's. The "adaptive" feature refers to the fact that the smoothing is applied based on the estimated channel error rate, and thus not used in error-free conditions.

Note, the 144-bit frame also includes a synchronization bit and an unused bit for future expansion of the CELP 1016 Federal Standard. Table 2.1 lists the bit allocation for all the parameters within a CELP encoded frame.

2.5 Sensitivity of Different CELP parameters

The CELP parameters also exhibit different degrees of sensitivity to transmission noise. In other words, channel errors in different CELP parameter bits produce different levels of perceptually annoying synthesis errors. Studies have been performed to measure the sensitivities of the different CELP parameters on both a bit by bit basis and a symbol by symbol basis [5], [30].

In Table 2.3, previously obtained results from bit by bit sensitivity tests are summarized [30]. Most entries in this table are of the form - a/b (**Symbol Label**)c. Here, a is the number of most significant bits, b is the total number of bits in that CELP parameter, and c is LSP number in the case of the LSP parameters or the subframe number in the case of all the other parameters. The **Symbol Labels** are summarized as follows:

LSP: Line Spectral Pair,

PD: Pitch Delay (Odd Subframes Only),

DD: Delta Delay (Even Subframes Only),

CBG: Codebook Gain,

PG: Pitch Gain,

CBI: Codebook Index.

These tests involved the corruption of individual bits of the CELP parameters, followed by measurements of the average segmental signal-to-noise, $Avg.\ SNR_{Segment}$, of the resulting CELP synthesized speech frames. The $SNR_{Segment}$ is computed for each CELP subframe as follows,

$$SNR_{Segment} = 10 \cdot log\left(\frac{E_{Signal}}{E_{Noise}}\right), \tag{2.1}$$

where, E_{Signal} is the energy of the CELP coder input waveform samples, $\{x_i\}$,

$$E_{Signal} = \frac{1}{N}\sum_{n=1}^{N} x_n^2, \tag{2.2}$$

and E_{Noise} is the energy of the difference between the input samples, $\{x_i\}$, and the CELP synthesized output waveform samples,$\{\hat{x}\}$,

$$E_{Noise} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{x}_n)^2, \tag{2.3}$$

where $N = 60$ since there are 60 samples per subframe. The average segmental signal-to-noise, $Avg.\ SNR_{Segment}$ is then calculated from equation (2.1) by averaging over all subframes. Note, that subframes with either zero signal or noise energy are excluded from this average.

In addition, subjective listening test are performed to determine which bits are most essential to understanding the output speech [30], [31].

Studies have also been performed to measure the overall CELP parameters sensitivity to noise. These studies are also based on the above mentioned criteria. The findings can be summarized as follows [30]:

- The most sensitive LSP's are LSP2, LSP3, LSP4 and LSP5.
- The second most sensitive LSP's are LSP1 and LSP6.
- The remainder of the LSP's can be ordered monotonically according to sensitivity – LSP7, LSP8, LSP9 and LSP10, where LSP10 is the least sensitive.

Most Sensitive	1/3 LSP1, 3/4 LSP2, 2/4 LSP3, 2/4 LSP4 2/4 LSP5, 1/3 LSP6, 1/3 LSP7 8/8 PD1, 8/8 PD3, 2/5 PG1, 2/5 PG2, 2/5 PG3, 2/5 PG4 2/5 CBG1, 2/5 CBG2, 2/5 CBG3, 2/5 CBG4
More Sensitive	1/3 LSP1, 1/4 LSP2, 2/4 LSP3, 1/4 LSP4 1/4 LSP5, 2/3 LSP6, 1/3 LSP7, 1/3 LSP8 6/6 DD1, 6/6 DD3, 3/5 PG1, 3/5 PG3, 3/5 PG3, 3/5 PG4 3/5 CBG1, 3/5 CBG3, 3/5 CBG3, 3/5 CBG4
Less Sensitive	1/3 LSP1, 1/4 LSP4, 1/4 LSP5, 1/3 LSP7, 2/3 LSP8, 3/3 LSP9, 3/3 LSP10
Least Sensitive	9/9 CBI1, 9/9 CBI2, 9/9 CBI3, 9/9 CBI4, 4/4 Hamming 1/1 Expansion Bit 1/1 Synchronization Bit

Table 2.3: Different sensitivity levels for the CELP parameter bits.

- The odd subframe pitch delays, PD1 and PD3, are the most sensitive of the excitation parameters.

- The codebook gains, CBG1, CBG2, CBG3, and CBG4, are the next most important of the excitation parameters, followed closely by the even subframe delta delays, DD2 and DD4.

- The pitch gains, PG1, PG2, PG3, and PG4, are the third most sensitive group of excitation parameters.

- The codebook indices, CBI1, CBI2, CBI3, and CBI4, are the least sensitive CELP parameters.

Note, that the other CELP frame bits (ie. Hamming code bits, synchronization bit, and the future bit) have no effect on the synthesis of CELP encoded speech.

Chapter 3

Source Optimized Channel Decoding of the LSP parameters

It was Shannon's Separation Principle [32] that first stated that source coding (or compression) and channel coding (or protection against channel errors) could be treated as separate entities without any loss in optimality. This is the major premise behind tandem source-channel coding, where the source compression is optimized for the given source description, and the error coding scheme is designed for a certain channel. However, this theorem is asymptotic in nature; it assumes infinitely long block sizes and unlimited encoder/decoder complexity. In actual systems, these long block sizes translate into long delays and increased complexity. If there are constraints placed on the acceptable delay/complexity of the system, as in today's wireless environments, it can be shown that joint source-channel techniques can outperform traditional tandem coding methodologies (e.g., [4], [9],[15]).

In [15] and [33], Hagenaur used a joint source-channel coding method that exploited the known characteristics of the source in the channel decoder's error protection scheme. In the paper by Alajaji *et. al.* [1], this "source-optimized channel decoding" approach was also used in transmitting the LSP parameters of Federal

Standard CELP 1016 over very noisy channels. We will be using the system developed in [1] in our studies.

3.1 Redundancy in the Line Spectral Pairs (LSP's)

The term "source-optimized channel decoding" in the case of the CELP LSP parameters simply refers to a decoding scheme that exploits the redundancy present in the encoded LSP parameters.

The best way to understand the reason for the redundancy existing in the CELP LSP parameters is to see how they are derived from the Linear Prediction filter [8],[10].

3.1.1 Deriving the LSP's from the LP coefficients

As previously mentioned, the LP analysis is done with an all-pole, minimum phase filter of order p as follows:

$$H(z) = \frac{1}{A_p(z)}. \tag{3.1}$$

In the determination of the LP filter coefficients, the inverse filter $A_p(z)$, is used

$$A_p(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_p z^{-p}, \tag{3.2}$$

where $\{a_i\}_{i=1,2,\ldots,p}$ are the LP coefficients.

It can be shown that $A_p(z)$ satisfies the following recursive relationship:

$$A_{p+1}(z) = A_p(z) - k_{p+1} z^{-(p+1)} A_p(z^{-1}), \tag{3.3}$$

$$\text{where } A_0(z) = 1,$$

and the $\{k_i\}_{i=1,2,\ldots,p}$ parameters are the PARCOR* coefficients [29]. Using equation (3.3) and the boundary conditions $k_{p+1} = 1$ and $k_{p+1} = -1$, which correspond,

*Another representation of the LP coefficients.

respectively, to the two extreme cases of the glottis being completely closed and completely opened†, the following $(p+1)^{th}$ order equations are derived:

$$P_{p+1}(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1}), \tag{3.4}$$

$$Q_{p+1}(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1}), \tag{3.5}$$

such that,

$$A_p(z) = \frac{P_{p+1}(z) \; + \; Q_{p+1}(z)}{2}. \tag{3.6}$$

Equations (3.4) and (3.5) can then be re-expressed as,

$$P_{p+1}(z) = (1 - z^{-1}) \prod_{i=2,4,\ldots,p} (1 - 2z^{-1} \cos \omega_i + z^{-2}), \tag{3.7}$$

$$Q_{p+1}(z) = (1 + z^{-1}) \prod_{i=1,3,\ldots,p-1} (1 - 2z^{-1} \cos \omega_i + z^{-2}), \tag{3.8}$$

with corresponding roots, $e^{j\omega_i}, i = 1, 2, ..., p$, where the parameters $\{\omega_i\}_{i=1,2,\ldots,p}$ are known as the *line spectral pairs*. Note, that it is assumed that p is an even integer and that $\omega_1 < \omega_3 < ... < \omega_{p-1}$ and $\omega_2 < \omega_4 < ... < \omega_p$. Also, $\omega_0 = 0$ and $\omega_{p+1} = \pi$ are the fixed roots of $P_{p+1}(z)$ and $Q_{p+1}(z)$, respectively.

Thus, the LSP parameters can be interpreted as the resonant frequencies of the vocal tract under the two extreme conditions of the glottis – opened and closed. Some important properties of the LSP parameters are summarized as follows:

- All the roots of $P_{p+1}(z)$ and $Q_{p+1}(z)$ lie on the unit circle.

- The roots of $P_{p+1}(z)$ and $Q_{p+1}(z)$ alternate on the unit circle, such that,

$$0 = \omega_0 < \omega_1 < ... < \omega_p < \omega_{p+1} = \pi. \tag{3.9}$$

It is clear from equation (3.9) that the LSP's within a frame are *ordered*, and thus, *highly correlated*. The "ideal" source encoder would accept the LSP parameters and

†The glottis is an opening in the vocal folds used in the vocal tract model [8].

output a stream of independent and identically distributed (i.i.d.) equiprobable bits. However, the CELP 1016 LSP scalar quantizer is not an "ideal" source encoder, as previous work has shown [10]. Thus, *residual redundancy* remains in the quantized LSP parameters. Through mathematical modeling and quantification, this residual redundancy is used in the channel decoder to protect against channel errors [1].

3.1.2 Definition of Residual Redundancy

Residual redundancy reflects the residual correlation (or memory) and nonuniformity left in the bitstream after it is passed through a source encoder. Intuitively, residual redundancy can be thought of as the ability to compress the output bitstream even further.

Mathematically, it can be understood in terms of the *entropy rate*, $H(\mathcal{X})$, of a stochastic process $\{X_i\}_{i=1,2,...}$ with finite alphabet $\mathcal{X}$. Where $H(\mathcal{X})$ is defined by,

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, ..., X_n), \tag{3.10}$$

when the limit exists, and,

$$\begin{aligned} H(X_1, X_2, ..., X_n) = - \sum_{x_1,...,x_n} & Pr(X_1 = x_1, ..., X_n = x_n) \\ & \cdot \, log_2 \; Pr(X_1 = x_1, ..., X_n = x_n). \end{aligned} \tag{3.11}$$

$H(\mathcal{X})$ represents the asymptotic per symbol entropy of the n random variables $\{X_i\}_{i=1,2,...,n}$, where n grows to infinity.

There is also an interesting property relating to the entropy rate of stationary stochastic processes, which is presented in the following theorem [7]:

Theorem 1: *For a stationary, stochastic process $\{X_i\}_{i=1,2,...}$ with finite alphabet $\mathcal{X}$, $H(\mathcal{X}) = \lim_{n\to\infty} H(X_n|X_{n-1}, X_{n-2}, .., X_1)$.*

If we now consider an i^{th} order *Markov chain* or *Markov process* $\{X_i\}_{i=1,2,...}$ with finite alphabet $\mathcal{X}$, then for $n > i$

$$\begin{aligned} & Pr(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, ..., X_1 = x_1) \\ = \; & Pr(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, ..., X_{n-i} = x_{n-i}). \end{aligned} \tag{3.12}$$

Combining this result with Theorem 1, we find that the entropy rate of a stationary, i^{th} order Markov process is given by

$$H(\mathcal{X}) = \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2}, .., X_{n-i}). \tag{3.13}$$

The total residual redundancy, ρ_T, of a stationary process defined as

$$\rho_T \triangleq log_2|\mathcal{X}| \; - \; H(\mathcal{X}), \tag{3.14}$$

where $|\mathcal{X}|$ is the cardinality (total number of elements) of the source, $\mathcal{X}$. The total redundancy, ρ_T, can be divided into two parts – the redundancy due to the non-unformity of the source and the redundancy due to memory in the source, ρ_D and ρ_M respectively:

$$\rho_D \triangleq log_2|\mathcal{X}| \; - H(X_1), \tag{3.15}$$

$$\rho_M \triangleq H(X_1) - H(\mathcal{X}), \tag{3.16}$$

hence,

$$\rho_T = \rho_D + \rho_M. \tag{3.17}$$

$H(X_1)$ is defined as the entropy of random variable X_1:

$$H(X_1) = - \sum_{x_1} Pr(X_1 = x_1) \cdot \log_2 Pr(X_1 = x_1), \tag{3.18}$$

where, intuitively, $H(X_1)$ represents the minimum number of bits needed to represent X_1.

Now that redundancy measures have been defined mathematically, a model can be developed to quantify the residual redundancy left in the CELP 1016 LSP parameters.

3.1.3 Modeling the LSP Parameters

Our approach to modeling the LSP parameters is the same as in [1]. As previously mentioned, in CELP 1016 the LSP parameters are quantized by either a 3-bit or 4-bit scalar quantizer. The quantized LSP's are guaranteed to be ordered by equation (3.9), such that $LSP1 < LSP2 < ... < LSP10$. In our model, we quantify the redundancy in the 3 most significant bits (MSB's) of each LSP, thus, ignoring the least significant bit in LSP's two through five.

Let the random process, $\{U_{i,j} : 1 \leq i \leq 10, j = 1, 2, ...\}$, represent the three most significant bits of the i^{th} quantized LSP in frame j. Thus, $\mathbf{U}_j = [U_{1,j}, U_{2,j}, ..., U_{10,j}]$ is a vector consisting of the 10 quantized LSP's (3 MSB's) in frame j.

If we assume that the process, $\{\mathbf{U}_j\}_{j=1,2,...}$, is stationary‡ , and use Theorem 1, then we can state that:

$$H(\mathcal{U}) = \lim_{j \to \infty} H(\mathbf{U}_j | \mathbf{U}_{j-1}, \mathbf{U}_{j-2}, .., \mathbf{U}_1), \tag{3.19}$$

where $H(\mathcal{U})$ represents the minimum number of bits per frame required to describe $\{\mathbf{U}_j\}_{j=1,2,...}$.

We now present two models§ for quantifying the redundancy present in the encoded LSP's:

- **Model A** assumes that LSP's in different frames are independent, and that the LSP's within a frame can be modeled by a 1^{st} order Markov process. This assumption is based on the ordered nature of the LSP's within a frame. This Markovian property can be stated between frames as

$$Pr(\mathbf{U}_j = \mathbf{u}_j | \mathbf{U}_{j-1} = \mathbf{u}_{j-1}, \mathbf{U}_{j-2} = \mathbf{u}_{j-2}, ..., \mathbf{U}_1 = \mathbf{u}_1)$$

‡No claim is made that the LSP blocks do form a stationary process. However, to the extent that the quantized LSP's can be *approximated* by a block stationary random process, the following calculations do indicate how much redundancy is present in those quantized LSP's.

§The only mathematical justification for these models is the correlation of the LSP parameters. Their appropriateness will be judged by the decoding algorithms based on that model.

$$= Pr(\mathbf{U}_j = \mathbf{u}_j), \tag{3.20}$$

and within a frame as

$$Pr(U_{i,j} = u_{i,j} | U_{i-1,j} = u_{i-1,j}, U_{i-2,j} = u_{i-2,j}, ..., U_{1,j} = u_{1,j})$$
$$= Pr(U_{i,j} = u_{i,j} | U_{i-1,j} = u_{i-1,j}) \tag{3.21}$$
$$\triangleq P_A^{(i)}(u_{i,j} | u_{i-1,j}), \tag{3.22}$$

for $i = 2, 3, ..., 10$ and $j = 1, 2, ...$ Note that for i=1, equation (3.22) becomes $P_A^{(1)}(u_{1,j})$.

- **Model B** assumes a 2^{nd} order Markov process: $U_{i,j}$ is independent of all previous LSP's conditioned on the immediately preceding LSP, $U_{i-1,j}$, and the corresponding LSP in the previous frame, $U_{i,j-1}$. This assumption is based not only on the ordered nature of the LSP's within a frame but also on the interpolation of LSP's from neighboring frames by the CELP adaptive smoother. More precisely, we have

$$Pr(U_{i,j} = u_{i,j} | \mathbf{U}_{j-1} = \mathbf{u}_{j-1}, ..., \mathbf{U}_1 = \mathbf{u}_1, U_{1,j} = u_{1,j}, ..., U_{i-1,j} = u_{i-1,j})$$
$$= Pr(U_{i,j} = u_{i,j} | U_{i-1,j} = u_{i-1,j}, U_{i,j-1} = u_{i,j-1}) \tag{3.23}$$
$$\triangleq P_B^{(i)}(u_{i,j} | u_{i-1,j}, u_{i,j-1}), \tag{3.24}$$

for $i = 2, 3, ..., 10$ and $j = 1, 2, ...$ Note that for i=1, equation (3.24) becomes $P_B^{(1)}(u_{1,j} | u_{1,j-1})$.

We next turn our attention to computing the residual redundancy in the encoded LSP's using both intraframe Markovity of Model A, and the intraframe and interframe Markovity of Model B.

First we note that the cardinality, $|\mathcal{U}|$, of the random process, $\{\mathbf{U}_j\}_{j=1,2,...}$, is simply 2^{30} since each frame is represented by 30 bits¶. Therefore, $log_2|\mathcal{U}| = 30$

¶Remember, we are only using the 3 MSB's of each LSP.

bits/frame. The total redundancy (per frame) can now be expressed as

$$\rho_T = 30 - H(\mathcal{U}) \text{ (bits/frame)}. \tag{3.25}$$

Using equations (3.19), (3.20) and (3.22), we can now derive an expression for $H(\mathcal{U})$ in the case of Model A:

$$\begin{aligned}
H(\mathcal{U}) &= \lim_{j\to\infty} H(\mathbf{U}_j|\mathbf{U}_{j-1}, \mathbf{U}_{j-2}, .., \mathbf{U}_1) && (3.26)\\
&= H(\mathbf{U}_j) && (3.27)\\
&= \sum_{i=1}^{10} H(U_{i,j}|U_{i-1,j}) && (3.28)\\
&= -\Bigg[\sum_{i=2}^{10} E_{P_A^{(i)}(u_{i,j},u_{i-1,j})}[log_2\ P_A^{(i)}(u_{i,j}|u_{i-1,j})] \\
&\quad + E_{P_A^{(1)}(u_{1,j})}[log_2\ P_A^{(1)}(u_{1,j})]\Bigg]. && (3.29)
\end{aligned}$$

Note that equation (3.27) follows from the fact that the entropy rate is independent of the frame j (due to block stationarity). The total residual redundancy in a frame exhibited by Model A becomes

$$\begin{aligned}
\rho_T &= 30 + E_{P_A^{(1)}(u_{1,j})}[log_2\ P_A^{(1)}(u_{1,j})] + \\
&\quad \sum_{i=2}^{10} E_{P_A^{(i)}(u_{i,j},u_{i-1,j})}[log_2\ P_A^{(i)}(u_{i,j}|u_{i-1,j})]. && (3.30)
\end{aligned}$$

Similarly using equations (3.19) and (3.24), an expression for $H(\mathcal{U})$ in the case of Model B can be derived:

$$\begin{aligned}
H(\mathcal{U}) &= \lim_{j\to\infty} H(\mathbf{U}_j|\mathbf{U}_{j-1}, \mathbf{U}_{j-2}, .., \mathbf{U}_1) && (3.31)\\
&= H(\mathbf{U}_j|\mathbf{U}_{j-1}) && (3.32)\\
&= \sum_{i=1}^{10} H(U_{i,j}|U_{i-1,j}, U_{i,j-1}) && (3.33)\\
&= -\Bigg[\sum_{i=2}^{10} E_{P_B^{(i)}(u_{i,j},u_{i-1,j},u_{i,j-1})}[log_2\ P_B^{(i)}(u_{i,j}|u_{i-1,j}, u_{i,j-1})] \\
&\quad + E_{P_B^{(1)}(u_{1,j},u_{1,j-1})}[log_2\ P_B^{(1)}(u_{1,j}|u_{1,j-1})]\Bigg]. && (3.34)
\end{aligned}$$

The total residual redundancy for Model B reduces to

$$\begin{aligned}\rho_T &= 30 + E_{P_B^{(1)}(u_{1,j},u_{1,j-1})}[log_2\ P_B^{(1)}(u_{1,j}|u_{1,j-1})] + \\ & \sum_{i=2}^{10} E_{P_B^{(i)}(u_{i,j},u_{i-1,j},u_{i,j-1})}[log_2\ P_B^{(i)}(u_{i,j}|u_{i-1,j},u_{i,j-1})]. \end{aligned} \tag{3.35}$$

To find the amount of redundancy due to non-uniformity, and the amount due to memory, ρ_D and ρ_M need to be computed. To compute these redundancy values we need H^*, the process entropy rate if the LSP's were *independent*:

$$H^* = \sum_{i=1}^{10} H(U_{i,j}) \tag{3.36}$$

$$= -\sum_{i=1}^{10} E_{Pr(u_{i,j})}[log_2\ Pr(u_{i,j})]. \tag{3.37}$$

Using (3.15) and (3.37) the redundancy due to non-uniformity, ρ_D, for both models becomes:

$$\rho_D = 30 + \sum_{i=1}^{10} E_{Pr(u_{i,j})}[log_2\ Pr(u_{i,j})]. \tag{3.38}$$

The redundancy due to memory, ρ_M, for Model A and B are:

$$\rho_M^{(A)} = H^* - H(\mathcal{U}) \tag{3.39}$$

$$\begin{aligned} &= -\sum_{i=1}^{10} E_{Pr(u_{i,j})}[log_2\ Pr(u_{i,j})] + \\ & E_{P_A^{(1)}(u_{1,j})}[log_2\ P_A^{(1)}(u_{1,j})] + \\ & \sum_{i=2}^{10} E_{P_A^{(i)}(u_{i,j},u_{i-1,j})}[log_2\ P_A^{(i)}(u_{i,j}|u_{i-1,j})], \end{aligned} \tag{3.40}$$

and

$$\rho_M^{(B)} = H^* - H(\mathcal{U}) \tag{3.41}$$

$$\begin{aligned} &= -\sum_{i=1}^{10} E_{Pr(u_{i,j})}[log_2\ Pr(u_{i,j})] + \\ & E_{P_B^{(1)}(u_{1,j},u_{1,j-1})}[log_2\ P_B^{(1)}(u_{1,j}|u_{1,j-1})] + \\ & \sum_{i=2}^{10} E_{P_B^{(i)}(u_{i,j},u_{i-1,j},u_{i,j-1})}[log_2\ P_B^{(i)}(u_{i,j}|u_{i-1,j},u_{i,j-1})], \end{aligned} \tag{3.42}$$

respectively.

A large training sequence (83,826 frames) from the TIMIT speech database [27] was input to the Federal Standard CELP 1016 vocoder. For every 30ms of speech, LPC analysis was performed to arrive at 10 quantized LSP's. The relative frequency of transitions between the values of 3 MSB's of each LSP were compiled to compute the Markov transition probabilities of both models. Then, using equations (3.30), (3.35), (3.38), (3.40) and (3.42), the different redundancy results were computed as shown in Table 3.1. Note, that values of ρ_D, ρ_M, and ρ_T for each of the 10 individual LSP's as well as the entire frame are shown.

LSP Redundancy	Model A			Model B		
	ρ_D	ρ_M	ρ_T	ρ_D	ρ_M	ρ_T
LSP1	0.6816	0.0000	0.6816	0.6816	0.2765	0.9581
LSP2	0.4805	0.4415	0.9219	0.4805	0.8469	1.3273
LSP3	0.7566	0.4425	1.1991	0.7566	0.7624	1.5190
LSP4	0.7093	0.4303	1.1396	0.7093	0.7529	1.4622
LSP5	0.3495	0.7184	1.0679	0.3495	0.8986	1.2481
LSP6	0.3585	0.6287	0.9872	0.3585	0.9367	1.2952
LSP7	0.6764	0.7575	1.4339	0.6764	0.8144	1.4908
LSP8	0.4511	0.3521	0.8032	0.4511	0.7990	1.2501
LSP9	0.2953	0.3840	0.6793	0.2953	0.6224	0.9177
LSP10	0.5160	0.4377	0.9537	0.5160	0.5007	1.0167
Frame Redundancy	5.2747	4.5927	9.8674	5.2747	7.2105	12.4852

Table 3.1: LSP Redundancies (in Bits/Frame) using 83,826 Frames of the TIMIT Speech Database.

Interpretation of results:

- Model A shows that accounting for only the correlation within a frame, 9.867 bits (ρ_T) of each 30 bit frame are redundant, with approximately 5.275 bits (ρ_D) of redundancy due to the non-uniform probability distribution of the LSP's, and 4.593 (ρ_M) bits of redundancy due to memory within a frame.

- Model B shows that accounting for not only the correlation within a frame, but also the correlation between frames, 12.485 bits (ρ_T) of each 30 bit frame are redundant. Of these redundant bits, approximately 5.275 bits (ρ_D) are due to the non-uniformity of the LSP's, and 7.211 (ρ_M) bits are due to memory within a frame and between frames.

Intuitively, these results suggest that each frame of CELP 1016 encoded speech can be further compressed by approximately 12.5 bits (using Model B). Thus, only 17.5 bits of "real" information are sent per frame. In essence, the CELP encoder is unintentionally acting as a rate 17.5/30 channel encoder by adding 12.5 bits of redundancy, which can then be exploited in the channel decoder to protect against errors. Methods for exploiting this redundancy via equal error protection (EEP) channel coding are discussed next.

3.2 An Equal Error Protection Scheme (EEP) for LSP Transmission

3.2.1 System Model

The diagram of the overall system used for EEP channel coding of the LSP's is shown in Figure 3.1.

A 32-state 3/4 rate convolutional code is used to protect *all* the 3-bit quantized LSP symbols, $\{\mathbf{u}_k\}$. The code has $d_{free} = 5$ and generator matrix [22]

$$\mathbf{G}(D) = \begin{bmatrix} 1+D & D & D & 1+D \\ D^2 & 1+D & 0 & 1+D+D^2 \\ 0 & D & 1+D^2 & 1+D^2 \end{bmatrix}. \tag{3.43}$$

Each symbol, $\mathbf{u}_k$, consists of the three MSB's of the quantized LSP. For each frame LSP1 is encoded first, followed by LSP2, and so on. Once LSP10 of a frame is encoded, then LSP1 of the next frame is encoded.

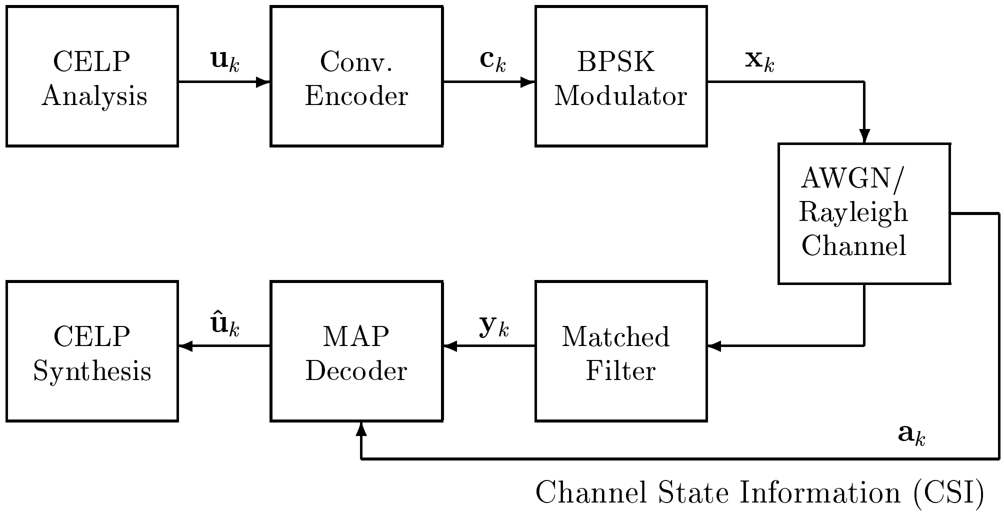

Figure 3.1: Block Diagram of the EEP Transmission System with MAP Decoding.

The rate is 3/4 so a 4-bit code symbol, $\mathbf{c}_k$, is produced in response to every three-bit LSP symbol, $\mathbf{u}_k$. The stream of code symbols, $\{\mathbf{c}_k\}$, is then BPSK modulated. The resulting real 4-tuple, $\mathbf{x}_k \in [-1, +1]^4$, is then sent over the channel according to:

$$\mathbf{y}_k = \mathbf{a}_k \mathbf{x}_k + \mathbf{n}_k, \tag{3.44}$$

where $\mathbf{a}_k$ and $\mathbf{y}_k$ are, respectively, the real 4-tuples corresponding to the channel fading coefficients and the output at time k. Note that $\mathbf{a}_k \mathbf{x}_k$ is the component-wise product of $\mathbf{a}_k$ and $\mathbf{x}_k$. The noise process, $\mathbf{n}_k$ is 4-tuple of zero-mean, independent Gaussian random variables with variance $N_0/2$.

The least significant bits of LSP's 2 through 5 are sent uncoded as a BPSK modulated real 4-tuple, and are corrupted as in equation (3.44) depending on the channel criteria.

The distribution of the fading coefficient $\mathbf{a}_k$ depends on the channel employed:

- For the purely additive white Gaussian noise (AWGN) channel, we assume $\mathbf{a}_k$ is the 4-dimensional all one vector.

- For the Rayleigh fading channel, we assume $\mathbf{a}_k$ consists of 4 independent random variables each with a Rayleigh distribution, where $E[a_i^2] = 1$. The variables are independent because we assume the channel is fully interleaved.

3.2.2 Maximum A-Posteriori (MAP) Decoding

The residual redundancy present in the LSP parameters can now be exploited in a convolutional decoder by adjusting the Viterbi algorithm decoding metric to incorporate the Markov transitional probabilities of our two models [1]. Maximum A-Posteriori decoding is optimal in the sense of minimizing the sequence error probability [2].

The MAP decision rule is to choose the code sequence, $\{\hat{\mathbf{c}}_k\}$ or $\{\hat{\mathbf{x}}_k\}$, which maximizes [22]

$$Pr(\mathbf{y}^K \mid \mathbf{x}^K)\; Pr(\mathbf{x}^K), \tag{3.45}$$

where, $\mathbf{y}^K \triangleq (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_K)$ is the corrupted received sequence, $\mathbf{x}^K \triangleq (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_K)$ is the sequence of the transmitted, modulated code symbols, and K is the sequence length (a multiple of 10). Note that each element, $\hat{\mathbf{x}}_k$, is a 4-tuple directly corresponding to the estimated LSP symbol, $\hat{\mathbf{u}}_k$.

For the AWGN/Rayleigh channel, the above metric reduces to choosing $\{\hat{\mathbf{x}}_k\}$ that minimizes

$$\sum_{k=1}^{K} \| \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \|^2 \; - \; N_0 \; ln \; Pr(\hat{\mathbf{x}}^K), \tag{3.46}$$

where the log_e of the above metric has been taken, and $\{\mathbf{a}_k\}$ is the sequence of Rayleigh fading coefficients‖, which we assume to be available at the decoder.

Now by applying the chain rule [28] on $ln \; Pr(\hat{\mathbf{x}}^K)$, we get

$$\begin{aligned} ln \; Pr(\hat{\mathbf{x}}^K) &= \sum_{k=1}^{K} ln \; Pr(\hat{\mathbf{x}}_k | \hat{\mathbf{x}}_{k-1}, \hat{\mathbf{x}}_{k-2}, ...) && (3.47) \\ &= \sum_{k=1}^{K} ln \; Pr(\hat{\mathbf{u}}_k | \hat{\mathbf{u}}_{k-1}, \hat{\mathbf{u}}_{k-2}, ...) && (3.48) \\ &\simeq \sum_{k=1}^{K} ln \; Pr(\dot{\mathbf{u}}_k | \dot{\mathbf{u}}_{k-1}, \dot{\mathbf{u}}_{k-2}, ...). && (3.49) \end{aligned}$$

Here equation (3.48) represents the universal distribution of the LSP symbols, which we have approximated with our sample distribution in equation (3.49).

Hence (3.46) becomes

$$\sum_{k=1}^{K} \| \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \|^2 \; - \; N_0 \; ln \; Pr(\mathbf{u}_k | \mathbf{u}_{k-1}, \mathbf{u}_{k-2}, ...). \tag{3.50}$$

Note that (3.50) can easily be implemented via a modified version of the Viterbi algorithm ([2]).

As in [1], we consider three soft decision decoding schemes based on 3.50,

- **ML:** Maximum likelihood decoding which chooses the code sequence $\{\hat{\mathbf{x}}_k\}$ that minimizes

$$\sum_{k=0}^{K} \| \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \|^2. \tag{3.51}$$

  Here, no redundancy in the LSP's is exploited in the decoding metric.

‖For AWGN, $\mathbf{a}_k = (1, 1, 1, 1)$ for all k.

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

$$\sum_{k=0}^{K} \| \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \|^2 \; - \; N_0 \; ln \; P_A^{([k \text{ mod } 10])}(\mathbf{u}_k|\mathbf{u}_{k-1}), \tag{3.52}$$

where $[k \text{ mod } 10]$ refers to the unique integer between 1 and 10. More explicitly, when $k \text{ mod } 10 = 0$ then $[k \text{ mod } 10] = 10$, otherwise $[k \text{ mod } 10] = k \text{ mod } 10$.

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

$$\sum_{k=0}^{K} \| \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \|^2 \; - \; N_0 \; ln \; P_B^{([k \text{ mod } 10])}(\mathbf{u}_k|\mathbf{u}_{k-1}, \mathbf{u}_{k-10}). \tag{3.53}$$

3.3 Simulation Results for EEP System

In the simulations that follow, the previously mentioned decoding schemes, based on the Viterbi algorithm, are used to decode the channel output sequence, $\{\mathbf{y}_k\}$.

The system deals with transmitted sequences of LSP's that correspond to thousands of frames of speech. A traditional sequence convolutional decoder would wait for the entire sequence to be received before decoding. The delay produced by Viterbi sequence decoding, as such, is not efficient. The practical implementation of the Viterbi decoder has a *path memory*, p, so that decoding can be accomplished with a much shorter delay. It has been shown that when $p \geq 5.8m$, where m is the maximal memory order (the length of the longest shift register in the encoder), the error due to truncation of the decoding process is negligible [11].

The path memory simply means that when a symbol y_k is received by the decoder, an estimate, z_{k-p}, is output by the decoder. The estimate, z_{k-p}, is the $(k-p)^{th}$ information block on the surviving path on the decoder trellis with the best metric [6].

In our system, a path memory of $p_{EEP} = 10 = 5m_{EEP}$**was used, which corresponds to 10 LSP symbols. Thus, upon receiving y_k, the decoder releases the estimate $\hat{u}_{k-10}$. As previously mentioned, $\hat{u}_{k-10}$ is chosen as the symbol on the path with the minimum cumulative metric up to time k. Thus, the overall decoding delay of our system is roughly one frame or 30ms, which is within acceptable limits for speech transmission.

In our LSP transmission scheme there is a possibility that the LSP's within a frame are not ordered. This is due to the fact that the LSB's of LSP 2 through 5 are not channel coded, and hence, have a considerable probability of being corrupted. If we obtain an unordered LSP vector, we simply reorder it to yield a proper output.

To evaluate the performance of the soft-decision decoding algorithms, MAP1 and MAP2 versus the ML algorithm, simulations were performed. Tests were run using a 4753-frame (around 2.2 minutes of speech) testing sequence from the TIMIT speech database [27]. Care was taken to ensure that no speaker was used in both the training and testing sequences. Thus the approach used here was to employ a single "universal" model – constructed from the very large training sequence – to decode all the speech samples.

Three performance criteria are used in evaluating the decoding algorithms:

- *Average Spectral Distortion*, which is the most commonly used distortion measure for the LSP's [19], is defined as

$$SD_{Avg} = \frac{1}{T}\sum_{j=1}^{T}\left[\int_{-\pi}^{\pi}(10 \cdot log_{10}S_j(\omega) - 10 \cdot log_{10}\hat{S}_j(\omega))^2\frac{d\omega}{2\pi}\right]\ (dB) \tag{3.54}$$

  where $S_j(\omega)$ and $\hat{S}_j(\omega)$ are the original and reconstructed short term spectra associated with frame j, and T is the total number of frames. A related criteria is the percentage of outliers. In our simulations we used percentage of outliers

**Our rate 3/4 coder has a maximal memory $m_{EE} = 2$.

greater than 4 dB, which is simply the fraction of frames with a spectral distortion value greater than 4 dB. Note, that the CELP LSP quantizer alone, when the channel is noiseless, produces an average spectral distortion of 1.50 dB with 0.08% of outliers greater than 4 dB. Intuitively, a SD_{Avg} of 1 dB or less is equivalent to perceptually transparent encoding [19].

- *Probability of symbol error*, P_s, is simply the percentage of LSP symbols in error after decoding. This value does not include any errors introduced by the LSB of LSP's 2 through 5 (i.e., only errors in the three MSB's of each LSP are taken into account).

- *Average Speech Distortion*, which is an average of seven different speech distortion measures of two different groups – cepstral measures and cosh measures. In Appendix A, we define the average speech distortion measure and all of its components. Note, that this measure is averaged over all subframes where subframes with either zero signal or noise energy are excluded from the average. Remark that when the channel is noiseless, the average speech distortion is 4.79 dB.

Tables 3.2 to 3.5 and Figures 3.2 to 3.7 show all the performance criteria for the different decoding schemes, ML, MAP1, and MAP2, over both AWGN and Rayleigh fading channels. E_b refers to the *energy per information bit*, so for a rate R code,

$$\frac{E_b}{N_0} = \frac{1}{R}\frac{E_s}{N_0}, \tag{3.55}$$

where, E_s is the *energy per symbol*. Note that for our simulations we did not use integer values for E_b/N_0. However, by using the above equation our values for E_b/N_0 correspond to integer values for E_s/N_0. This was done to facilitate comparisons with the results in [1].

Observations Regarding Results:

- Tables 3.2 and 3.4 describe the performance of the equal error protection scheme over the AWGN channel. We see that at an E_s/N_0 of as low as 1 dB (E_b/N_0 = 2.12 dB[††]), both MAP schemes perform almost as if the channel was noiseless with a average spectral distortion of 1.61 dB. Note, the best possible value of average SD is 1.51 dB. Similarly, both MAP decoders are much closer to the optimal speech distortion value of 4.79 dB at an E_b/N_0 of 1.12 dB. More specifically, the MAP decoders are at approximately 5 dB compared to the ML decoder's speech distortion of 7.2 dB at an E_b/N_0 of 1.12 dB.

- At low SNR's, both MAP schemes have symbol error rates of around 1% compared to a symbol error rate of 10% to 30% for the ML decoder.

- Tables 3.3 and 3.5 contain results based on the above system over the Rayleigh fading channel. We see that with MAP2, we reach a symbol error rate of less than 1% at an E_s/N_0 as low as 3 dB. The corresponding ML decoder symbol error rate at this SNR is over 38%. Also, at this channel SNR, both average spectral distortion values for the MAP decoding schemes are well below 2 dB, whereas the ML decoder's spectral distortion is 5.71 dB.

††We assume a rate $R = 34/44$, which includes the uncoded LSB's of LSP 2 through 5.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-3.88 dB	11.11	8.41	8.48
-2.88 dB	10.96	7.73	7.44
-1.88 dB	10.63	6.83	6.41
-0.88 dB	10.02	5.90	5.53
0.12 dB	8.96	5.28	5.10
1.12 dB	7.20	4.96	4.90
2.12 dB	5.56	4.83	4.83

Table 3.2: Average Speech Distortion for Rate 3/4 Equal Error Protection of LSP parameters over the AWGN channel.

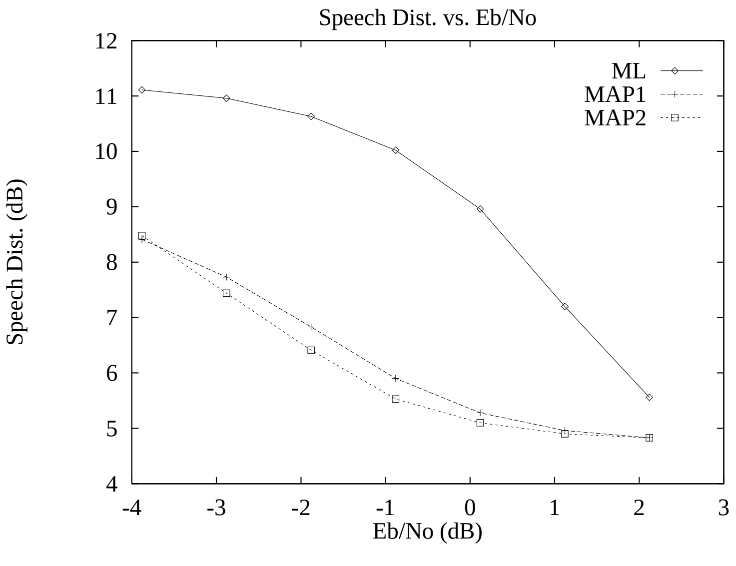

Figure 3.2: Average Speech Distortion for Different Decoding Schemes of LSP parameters over the AWGN Channel.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-2.88 dB	11.24	8.42	8.55
-1.88 dB	11.05	7.86	7.74
-0.88 dB	10.82	7.20	6.86
0.12 dB	10.56	6.54	6.08
1.12 dB	10.16	5.86	5.46
2.12 dB	9.62	5.35	5.14
3.12 dB	8.66	5.09	4.97
4.12 dB	7.40	4.92	4.89
5.12 dB	6.24	4.86	4.84
6.12 dB	5.46	4.83	4.84

Table 3.3: Average Speech Distortion of Rate 3/4 Equal Error Protection of LSP parameters over the Rayleigh Fading channel.

–	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.88 dB	10.46 (99.60%)	6.73 (69.58%)	6.46 (65.68%)	86.19%	46.12%	44.76%
-2.88 dB	10.28 (99.09%)	5.84 (58.56%)	5.27 (50.99%)	84.87%	37.05%	33.21%
-1.88 dB	9.91 (97.64%)	4.65 (43.14%)	3.96 (33.85%)	82.37%	25.78%	20.77%
-0.88 dB	9.29 (93.78%)	3.33 (25.46%)	2.74 (17.04%)	76.38%	14.40%	9.84%
0.12 dB	7.89 (80.69%)	2.36 (10.98%)	2.07 (6.90%)	61.96%	5.93%	3.71%
1.12 dB	5.42 (53.26%)	1.84 (3.75%)	1.73 (2.21%)	36.20%	1.83%	1.07%
2.12 dB	2.91 (19.76%)	1.62 (0.80%)	1.61 (0.65%)	11.49%	0.33%	0.27%

Table 3.4: Average Spectral Distortion and Symbol Error Rate for Rate 3/4 Equal Error Protection of LSP parameters over the AWGN channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-2.88 dB	10.58 (99.49%)	6.77 (69.21%)	6.57 (67.00%)	86.37 %	46.64 %	46.04 %
-1.88 dB	10.45 (99.30%)	6.05 (60.67%)	5.62 (55.52%)	85.51 %	39.50 %	37.06 %
-0.88 dB	10.21 (98.86%)	5.16 (49.58%)	4.54 (41.17%)	84.44 %	30.71 %	26.52 %
0.12 dB	9.90 (97.58%)	4.21 (36.95%)	3.51 (27.17%)	82.02 %	21.73 %	16.67 %
1.12 dB	9.45 (94.71%)	3.27 (24.31%)	2.66 (15.44%)	77.61 %	13.44 %	8.79 %
2.12 dB	8.76 (88.69%)	2.48 (13.47%)	2.15 (8.08%)	70.41 %	6.99 %	4.33 %
3.12 dB	7.49 (76.50%)	2.05 (6.90%)	1.87 (3.87%)	56.97 %	3.43 %	2.05 %
4.12 dB	5.71 (56.54%)	1.77 (2.63%)	1.73 (1.91%)	38.71 %	1.28 %	0.94 %
5.12 dB	3.93 (34.59%)	1.66 (1.18%)	1.64 (0.80%)	21.26 %	0.50 %	0.35 %
6.12 dB	2.67 (16.97%)	1.60 (0.44%)	1.61 (0.48%)	9.55 %	0.15 %	0.23 %

Table 3.5: Average Spectral Distortion and Symbol Error Rate for Rate 3/4 Equal Error Protection of LSP parameters over the Rayleigh Fading channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

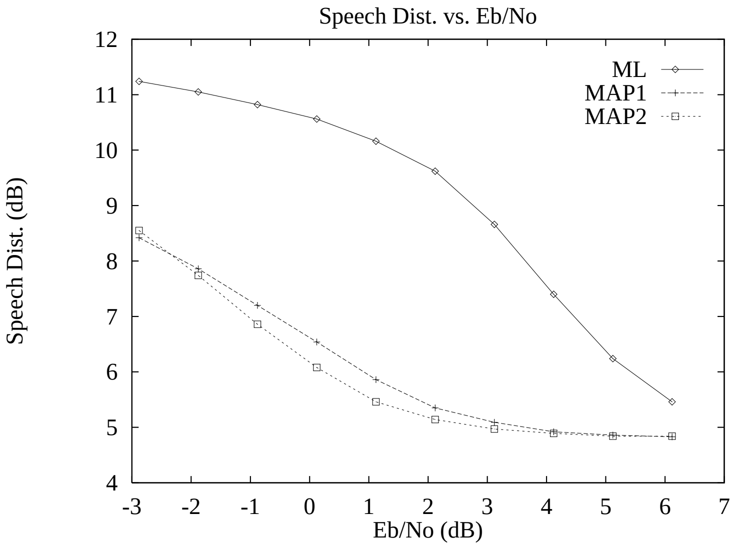

Figure 3.3: Average Speech Distortion for Different Decoding Schemes of LSP parameters over the Rayleigh Fading Channel.

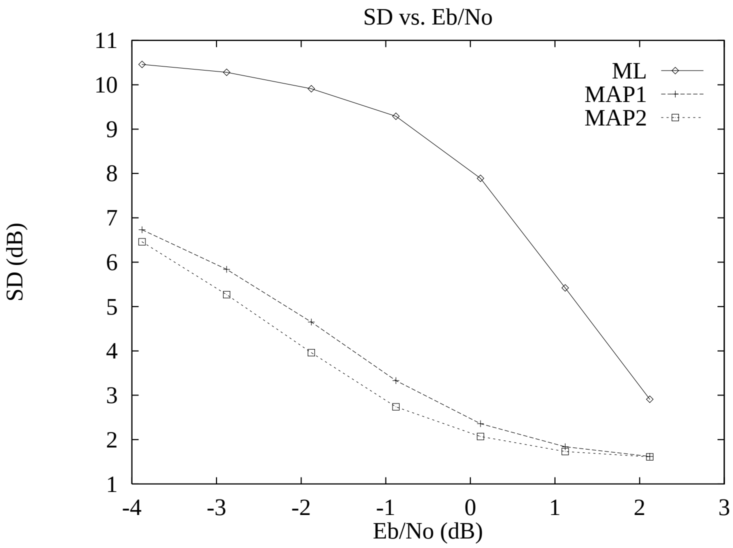

Figure 3.4: Average Spectral Distortion for Different Decoding Schemes of LSP parameters over the AWGN Channel.

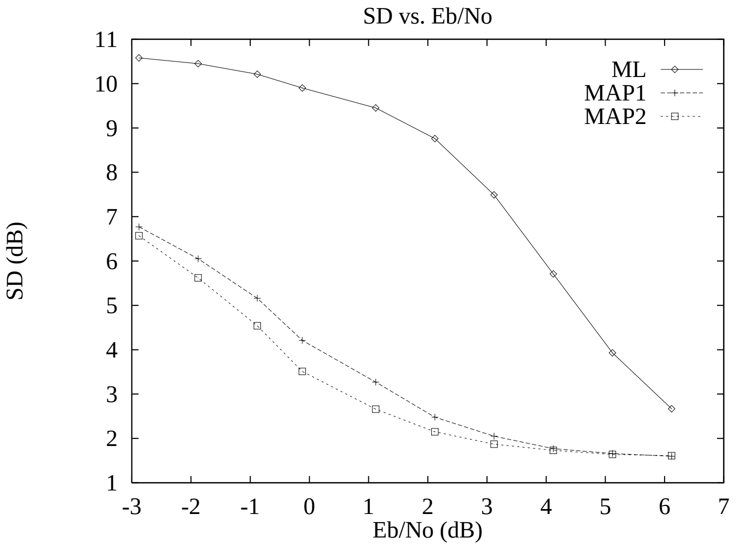

Figure 3.5: Average Spectral Distortion for Different Decoding Schemes of LSP parameters over the Rayleigh Fading Channel.

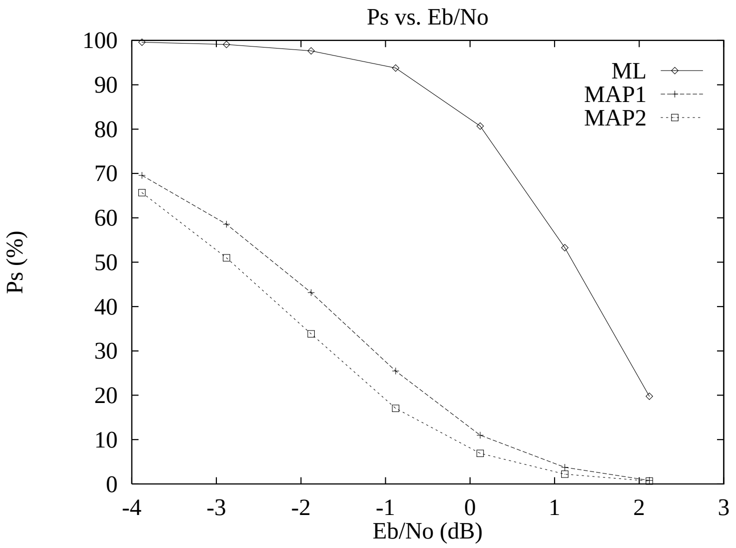

Figure 3.6: Symbol Error Rate for Different Decoding Schemes of LSP parameters over the AWGN Channel.

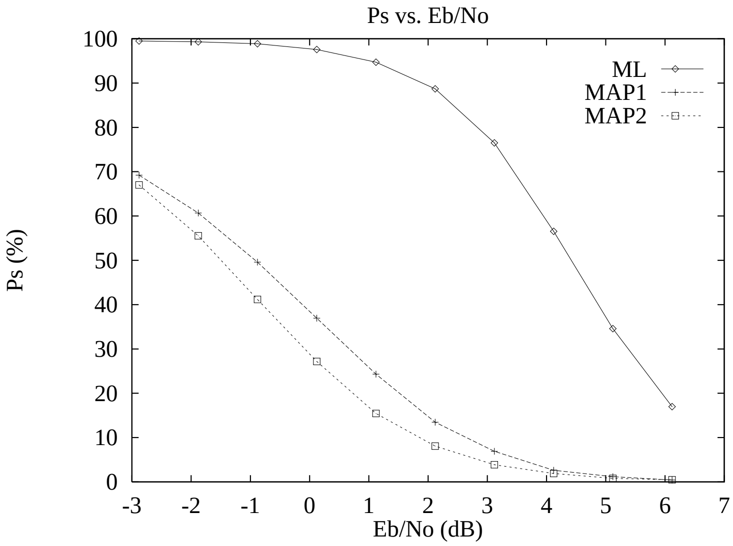

Figure 3.7: Symbol Error Rate for Different Decoding Schemes of LSP parameters over the Rayleigh Fading Channel.

Chapter 4

Source Optimized Channel Decoding and Unequal Error Protection (UEP) of the LSP parameters

As we already mentioned in Section 2.5, different LSP parameters exhibit different degrees of sensitivity to channel noise. Thus, it is paramount to protect those LSP's that are more important to the proper reconstruction of the sampled speech. Note that the way we modelled the redundancy in the LSP's will have an effect on their importance. More specifically, through MAP decoding, LSP1 will have an effect on the proper decoding of LSP2, which will in turn have an effect on the proper decoding of LSP3, and so on. Thus, in our simulations we have modified the findings of Section 2.5* and assumed that the lower LSP's are more important than the higher ones, with their corresponding MSB's holding greater significance than their less significant bits.

*In Section 2.5, it was established that LSP's 2 through 5 were the most important followed by LSP1 and LSP6.

4.1 Rate Compatible Punctured Convolutional (RCPC) Codes

We employ unequal error protection (UEP) schemes to allow different levels of protection for different LSP symbols. Our UEP systems consists of families of punctured convolutional codes [6] known as rate compatible punctured convolutional (RCPC) codes [14] that provide different levels of protection for the various LSP's, depending on their importance.

Punctured convolution codes were first suggested by Cain *et. al.* [6] as a means of attaining higher rate $R = k/n$ convolutional codes from lower rate $R = 1/n$ codes. The primary reason for these codes was the need to decrease the decoder complexity of higher rate codes. For a Viterbi implementation of rate $R = k/n$ code, 2^k comparisons must be made at each trellis node. It can easily be seen that as the memory of the high rate code increases, the decoder complexity becomes impractical for many applications. Thus, punctured convolutional codes were introduced to solve this problem. Punctured codes are generally based on rate $R = 1/n$ convolutional codes, and thus only require 2 comparisons at each trellis node.

Punctured codes can be attained by periodically perforating the output of low-rate convolutional codes (or mother codes), through a puncturing matrix, to obtain higher rate codes. 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 $\delta \in [1, (n-1)P]$ [20]. 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 zero's, which specify the punctured (or not transmitted) output bits, and ones, 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 $\mathbf{G}(D) = [1{+}D^2,\ 1{+}D{+}D^2]$. For puncturing period, $P = 2$, and puncturing matrix

$$\mathbf{A}(\delta)_{2\text{x}2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{4.1}$$

This puncturing rule essentially means delete every third output bit. The punctured trellis of the above code can be seen in Figure 4.1. 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

$$\mathbf{G}(D) = \begin{bmatrix} 1+D & 1+D & 1 \\ 0 & D & 1+D \end{bmatrix}. \tag{4.2}$$

The non-punctured trellis of a rate 2/3 code given by (4.2) is seen in Figure 4.2. 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 trivial, 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 simply 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

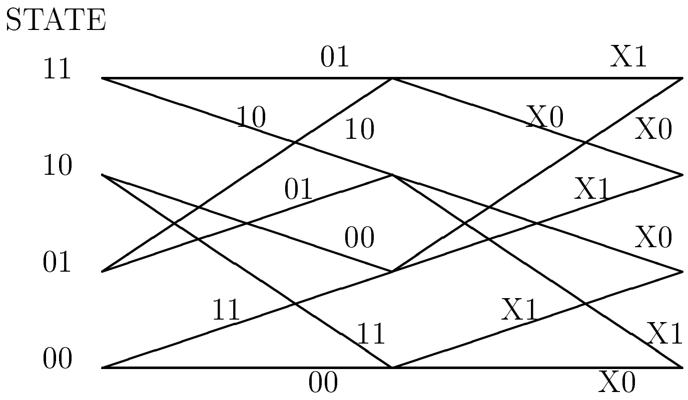

Figure 4.1: Trellis for rate $R = 1/2$, constraint length 3, period $P = 2$ punctured convolutional code.

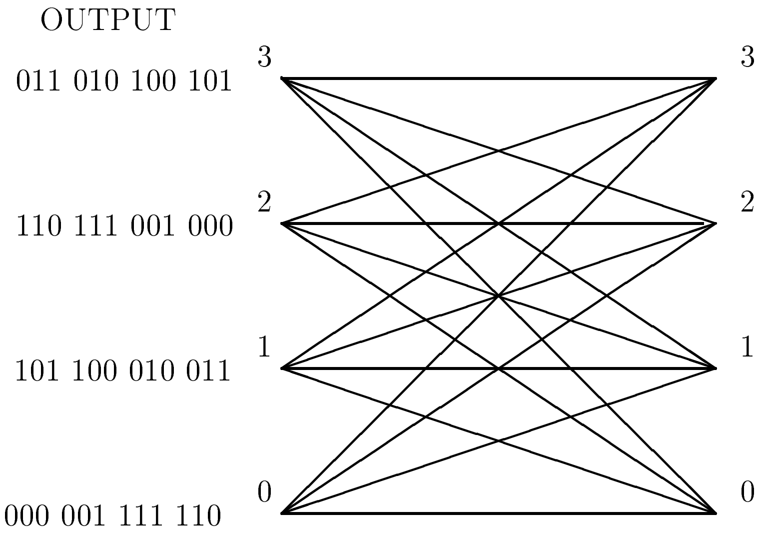

Figure 4.2: Trellis for rate $R = 2/3$, constraint length 3, non-punctured convolutional code.

a rate $R = 1/n$ mother code with period P, and

$$\mathbf{A}(\delta) = \begin{bmatrix} a_{11}(\delta) & \dots & a_{1P}(\delta) \\ \vdots & \ddots & \vdots \\ a_{n1}(\delta) & \dots & a_{nP}(\delta) \end{bmatrix} \tag{4.3}$$

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:

$$\text{If } a_{ij}(\delta) = 1 \text{ then } a_{ij}(\epsilon) = 1 \text{ for all } \epsilon \geq \delta \geq 1. \tag{4.4}$$

In other words, the puncturing matrix for the lower rate code, $\mathbf{A}(\epsilon)$ contains all the ones of the puncturing matrix for the higher rate code, $\mathbf{A}(\delta)$. The above condition simply gaurantees 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. In our case, we apply lower rate codes to the lower LSP's, which have been shown to be more paramount to proper speech reconstruction. Note that we cannot change our LSP order due to the Markov structure of both Model A and B. However, the original order of the LSP's is already in order of decreasing importance.

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 4.1 all places where an 'X' occurs is replaced by a zero in the calculation of the Viterbi decoding metric. 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 [21].

For a given puncturing matrix, $\mathbf{A}(\delta)$, within an RCPC code family the bound on bit error probability can be formulated as in the case of a traditional convolutional code. For instance, for a rate $P/(P+\delta)$ code the upper bound on bit error probability, P_B [13], is given by

$$P_B \leq \frac{1}{P} \sum_{d=d_{free}}^{\infty} c_d Pr(e_d), \tag{4.5}$$

where d_{free} refers to the free distance of the punctured code, c_d is the total number of error bits produced by incorrect error paths of Hamming weight d, and $Pr(e_d)$ is the probability of picking an incorrect error path in the Viterbi decoding process. For our BPSK modulation scheme and AWGN channel it can be shown that

$$Pr(e_d) = Q\left(\sqrt{2d\left(\frac{P}{(P+\delta)}\right)\frac{E_b}{N_0}}\right), \tag{4.6}$$

where

$$Q(y) \triangleq \int_y^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \tag{4.7}$$

and E_b/N_0 is the average bit energy to noise power spectral density ratio.

The "good codes generate good codes" approach [13] [20] is taken in developing families of RCPC codes. This entails using the maximal free distance code as the mother code. An exhaustive search is then performed to find the puncturing matrices that minimize the bit error probability as given by equation (4.5). In these searches the length of error paths, d, is not taken to infinity, instead a practical limit is chosen. In general, these punctured codes are not as good as the best known regular codes of the same rate.

No known work has been performed on finding the error bounds on an overall family of RCPC codes, which would include the effect of transitional phases on error performance.

4.2 An Unequal Error Protection (UEP) Scheme for LSP Transmission

4.2.1 System Model

The diagram of the overall system proposed for UEP channel coding of the LSP's is shown in Figure 4.3.

Three different families of RCPC codes are used. The first has a 32-state mother code of rate 1/2. Details on the generator matrix and corresponding puncturing matrices for puncturing periods 2 through 7 can be found in [20] or [21]. The next two families of RCPC codes, developed by Hagenaur [14], are based on a rate 1/3 32-state mother code and a rate 1/4 16-state code, respectively, both with period $P = 8$.

Each symbol, $\mathbf{u}_k$, consists of the three MSB's of the quantized LSP. Note that all our mother codes are now of base rate $1/n$. Thus, for each frame LSP1 is encoded first, starting with the MSB and proceeding to the LSB, followed by LSP2's MSB, and so on. Once the LSB of LSP10 of a frame is encoded, then the MSB of LSP1 of the next frame is encoded. In other words, each symbol, $\mathbf{u}_k$, is encoded one bit at a time.

The rate is $1/n$ so three n-bit code symbols, $\mathbf{c}_{k,i}$, (where $1 \leq i \leq 3$) are produced in response to each three-bit LSP symbol, $\mathbf{u}_k$. In other words the set $\mathbf{c}_k = \{\mathbf{c}_{k,1}, \mathbf{c}_{k,2}, \mathbf{c}_{k,3}\}$ corresponds to the three bit LSP symbol $\mathbf{u}_k$. The stream of code symbols, $\mathbf{c}_{k,i}$, is then punctured according to the RCPC family period, P, and puncturing matrices, $\mathbf{A}(\delta_i)$, where $1 \leq \delta_i \leq (n-1)P$.

The unpunctured bits in each n-bit codeword are then sent to a BPSK modulator,

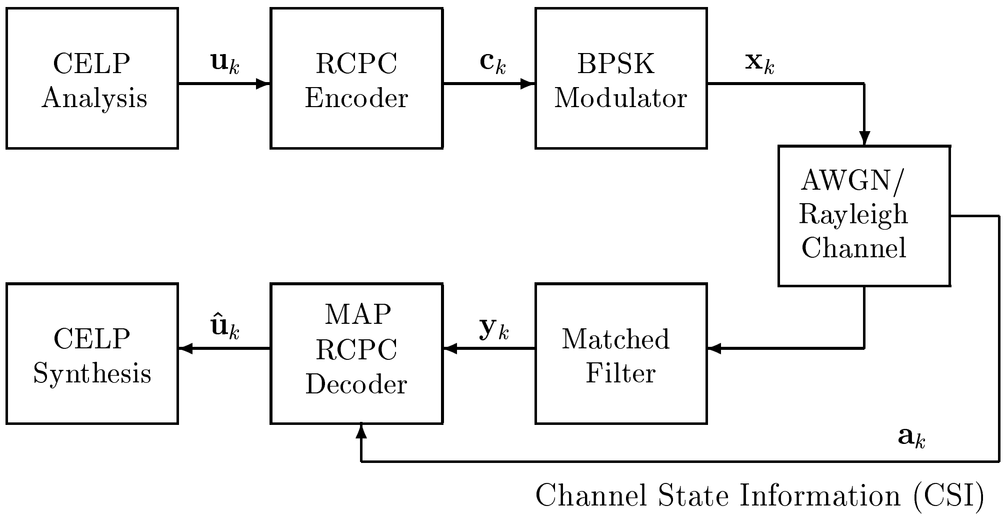

Figure 4.3: Block Diagram of the UEP System with MAP RCPC Decoder.

which produces real m-tuples, $\mathbf{x}_{k,i} \in [+1, -1]^m$, where $1 \leq m \leq n^{\dagger}$. The channel output, $\mathbf{y}_{k,i}$, is then produced according to

$$\mathbf{y}_{k,i} = \mathbf{a}_{k,i}\mathbf{x}_{k,i} + \mathbf{n}_{k,i}, \tag{4.8}$$

where $\mathbf{a}_{k,i}$ and $\mathbf{y}_{k,i}$ are, respectively, the real m-tuples $(1 \leq m \leq n)$ corresponding to the channel fading coefficients and the output at time $k+i$. As before, our noise process, $\mathbf{n}_{k,i}$ is m-tuple of zero-mean, independent Gaussian random variables with variance $N_0/2$.

Once again, the least significant bits of LSP's 2 through 5 for each frame are sent as a BPSK modulated real 4-tuple, and are corrupted as in equation (4.8) depending on the channel criterion. The channel criterion is either:

- The AWGN channel, where $\mathbf{a}_{k,i}$ is the m-dimensional all one vector, or
- The fully interleaved Rayleigh fading channel, where $\mathbf{a}_{k,i}$ consists of m independent random variables each with a Rayleigh distribution, with $E[a_i^2] = 1$.

4.2.2 Modified Viterbi Decoding

To use our previously derived MAP metric with this UEP scheme, we should use the transitional probabilities of the LSP bits, since we are transmitting one bit at a time. However, redundancy models based on LSP bits were 30% less redundant in bits per frame than those based on LSP's as three-bit symbols. Thus, we have chosen to keep our redundancy models based on the transitional probabilities of three-bit LSP symbols, $\mathbf{u}_k$. This shows the need for a convolutional coder of rate $3/n$ where each 3-bit LSP produces a corresponding codeword of n bits. Thus, we must modify our Viterbi algorithm so that we can use our redundancy models with our rate $R = 1/n$ convolutional coders. This modification consists of simply computing the corresponding

†In punctured convolutional codes, at least one bit of each codeword is always sent.

Viterbi metric every three trellis steps instead of every one. This then corresponds to 3 LSP bits or one full LSP symbol, so our MAP metrics can now be employed in the decoding phase. Thus, our new general metric for the AWGN/Rayleigh fading channel becomes

$$\sum_{k=1}^{K}\sum_{i=1}^{3}\left[\| \mathbf{y}_{k,i} - \mathbf{a}_{k,i}\hat{\mathbf{x}}_{k,i} \|^2\right] \; - \; N_0 \; ln \; Pr(\mathbf{u}_k|\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, ...), \tag{4.9}$$

where K is the length of the LSP symbol sequence, $\{\mathbf{u}_k\}$ (a multiple of 10).

The result of this modification is simply an increase in the UEP decoder complexity in that more comparisons are made at each trellis node. Originally, for a rate $1/n$ decoder with m memory elements, 2 comparisons are made at each of the 2^m nodes for each trellis step. In the modified decoding scheme, 8 comparisons (2^3) are made at each of the 2^m trellis nodes for each trellis steps. Where in the modified Viterbi algorithm a trellis step, consists of three trellis steps in the regular Viterbi algorithm. Thus, if our UEP mother code has 32 states (recall our 32 state rate 3/4 EEP convolutional coder) then it will have an equivalent decoding complexity as our EEP scheme.

The three soft decision decoding schemes based on the above modified Viterbi metric become:

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

$$\sum_{k=1}^{K}\sum_{i=1}^{3} \| \mathbf{y}_{k,i} - \mathbf{a}_{k,i}\hat{\mathbf{x}}_{k,i} \|^2. \tag{4.10}$$

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

$$\sum_{k=1}^{K}\sum_{i=1}^{3}(\| \mathbf{y}_{k,i} - \mathbf{a}_{k,i}\hat{\mathbf{x}}_{k,i} \|^2) \; - \; N_0 \; ln \; P_A^{([k \bmod 10])}(\mathbf{u}_k|\mathbf{u}_{k-1}), \tag{4.11}$$

  where $[k \bmod 10]$ refers to the unique integer between 1 and 10.

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

$$\sum_{k=1}^{K}\sum_{i=1}^{3}(\| \mathbf{y}_{k,i} - \mathbf{a}_{k,i}\hat{\mathbf{x}}_{k,i} \|^2) \; - \; N_0 \; ln \; P_B^{([k \bmod 10])}(\mathbf{u}_k|\mathbf{u}_{k-1}, \mathbf{u}_{k-10}). \tag{4.12}$$

4.2.3 MAP Decoding of Uncoded LSP Symbols

In our UEP simulations, we also consider transmitting LSP symbols with no channel coding. However, in this case we would still like to exploit the source characteristics, in the decoding of the unprotected LSP symbols.

For every LSP symbol, $\{\mathbf{u}_k\}$, there are three transmitted code symbols, $\{\mathbf{c}_{k,1}, \mathbf{c}_{k,2}, \mathbf{c}_{k,3}\}$, which then form the code symbol sequence, $\{\mathbf{c}_{k,i}\}$. If the LSP symbol is uncoded, the $\{\mathbf{c}_{k,i}\}$ corresponds to the individual bits of the LSP symbols, from most to least significant. This code symbol sequence is then modulated to form a sequence of 1-tuples, $\{x_{k,i}\}$, where $x_{k,i} \in [+1, -1]$, which are sent according to the channel criteria and equation (4.8) to get a corrupted received sequence $\{y_{k,i}\}$. Note also that $\{a_{k,i}\}$ and $\{n_{k,i}\}$ are now sequences of 1-tuples.

In UEP schemes where ML decoding is used with some of the LSP symbols sent uncoded, hard decision (HD) threshold decoding is employed on the uncoded LSP's. We use the following decoding rule for HD decoding

$$\hat{x}_{k,i} = \begin{cases} 0, & \text{if } y_{k,i} < 0 \\ 1, & \text{if } y_{k,i} \geq 0 \end{cases}, \tag{4.13}$$

and $\hat{x}_{k,i}$ is the estimated symbol with $y_{k,i}$ representing the symbol, $x_{k,i}$, after it had been corrupted by noise.

For our MAP1 and MAP2 systems, if uncoded LSP symbols are transmitted, we use similar metrics as in (4.11) and (4.12) with a sequence detection method. Thus we must chose the sequence of uncoded LSP symbol bits $\{\hat{x}_{k,i}\}$ that minimizes,

$$\sum_{k=1}^{K}\sum_{i=1}^{3}(\| y_{k,i} - a_{k,i}\hat{x}_{k,i} \|^2) \; - \; N_0 \; ln \; P_A^{([k \bmod 10])}(\mathbf{u}_k|\mathbf{u}_{k-1}), \tag{4.14}$$

for MAP1 decoding, and,

$$\sum_{k=1}^{K}\sum_{i=1}^{3}(\parallel y_{k,i} - a_{k,i}\hat{x}_{k,i} \parallel^2) \ - \ N_0 \ ln \ P_B^{([k \bmod 10])}(\mathbf{u}_k|\mathbf{u}_{k-1}, \mathbf{u}_{k-10}), \tag{4.15}$$

for MAP2 decoding.

To use the above MAP metrics in decoding, we employ an 8-state fully-connected trellis, where each state corresponds to a possible LSP symbol. Thus, it is possible to go from one state to any other state. The state of the trellis and its corresponding input are equivalent. Thus, each node on a path along this trellis represents an estimated unprotected LSP symbol, $\hat{u}_k$.

We now perform simulations to compute the results of completely uncoded transmission of the LSP parameters. Once again the 4753 frame testing sequence is used in conjunction with the hard decision (HD) and two soft decision (MAP1 and MAP2) decoding schemes. We have normalized the E_b/N_0 using equation (3.55), where $R = 34/34 = 1$ for uncoded transmission.

Note, decoding of uncoded LSP's is done frame-by-frame not by sequence as was done in the EEP simulations of the previous chapter. This will mean that in UEP systems where some LSP's are left unprotected, a parallel decoder design must be employed. However, the delay introduced by the 8-state trellis decoder of unprotected LSP's will also be 1 frame.

The results of our simulations are summarized in Tables 4.1 to 4.4. Figures 4.4 to 4.9 show the performance - in terms of the average spectral distortion - of our uncoded system versus the EEP channel coding system of the previous chapter for all three decoding schemes over AWGN and Rayleigh channels. Note that the graphical results for symbol error rate (P_s) and average speech distortion are not shown; they are similar in nature to those of the average spectral distortion.

Observations Regarding Results:

- Figure 4.4 shows that the completely uncoded system performs better than the coded one for low values of E_b/N_0. This is because we are designing our systems for very noisy channel conditions. One reason for this is that at low channel SNR's the error propagation effects in the convolutional decoder causes the EEP system to perform worse than the simple hard-decision uncoded system. The crossover point, when EEP starts performing better than uncoded is around $E_b/N_0 = 2$ dB for ML decoding over the AWGN channel.

- For MAP1 decoding over the AWGN channel, the EEP system starts to perform better than the uncoded system, at a crossover point of around $E_b/N_0 = -1$ dB (c.f, Figure 4.5). Similarly, it can be seen in Figure 4.6 that the crossover point for MAP2 decoding is even lower at $E_b/N_0 = -1.5$ dB.

- Figures 4.7 to 4.9 show the same results for the Rayleigh channel. Note that the performance of the EEP system compared to the uncoded system is similar to the AWGN channel. However, the crossover points are all at a higher E_b/N_0 since the Rayeligh fading channel suffers more distortion (due to fading) than the AWGN channel.

4.3 Simulation Results for UEP Systems

The same 4753-frame (2.2 minute) testing sequence was used in all the UEP simulations. The performance criteria for these systems remains the same as the EEP simulations. In other words, the average spectral distortion, average speech distortion and probability of symbol error are used to judge the performance of these systems.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-3.88 dB	8.14	6.66	6.60
-2.88 dB	7.76	6.28	6.21
-1.88 dB	7.34	5.98	5.90
-0.88 dB	6.88	5.69	5.62
0.12 dB	6.42	5.45	5.37
1.12 dB	6.00	5.26	5.19
2.12 dB	5.63	5.09	5.04

Table 4.1: Average Speech Distortion for Uncoded LSP parameters over the AWGN channel.

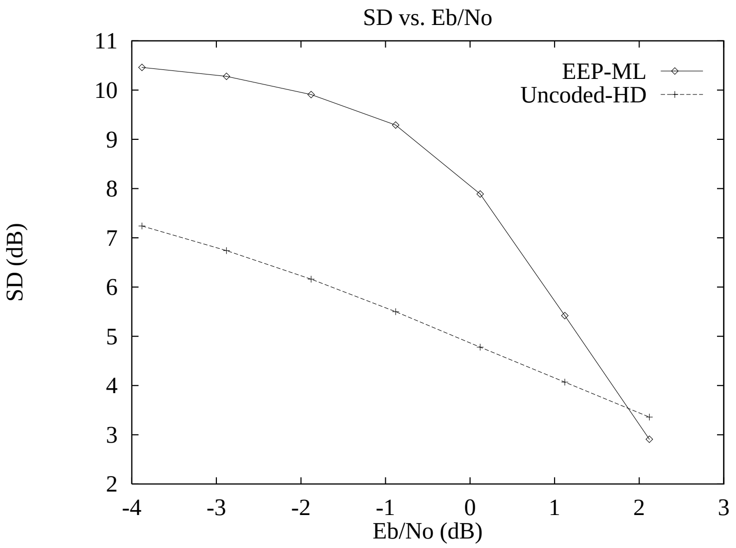

Figure 4.4: Spectral Distortion for Uncoded HD vs EEP ML Decoding Schemes of LSP parameters over the AWGN Channel.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-2.88 dB	8.50	6.74	6.70
-1.88 dB	8.18	6.46	6.40
-0.88 dB	7.87	6.21	6.09
0.12 dB	7.57	5.98	5.86
1.12 dB	7.29	5.79	5.67
2.12 dB	6.99	5.61	5.51
3.12 dB	6.70	5.46	5.38
4.12 dB	6.43	5.34	5.26
5.12 dB	6.19	5.25	5.17
6.12 dB	5.97	5.16	5.09

Table 4.2: Average Speech Distortion for Uncoded LSP parameters over the Rayleigh Fading channel.

–	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.88 dB	7.24 (93.24%)	4.83 (56.09%)	4.61 (52.25%)	45.17 %	30.01 %	29.38 %
-2.88 dB	6.74 (89.06%)	4.30 (47.02%)	4.07 (42.80%)	39.36 %	24.80 %	24.07 %
-1.88 dB	6.16 (82.87%)	3.83 (38.57%)	3.62 (33.99%)	33.20 %	20.08 %	19.19 %
-0.88 dB	5.50 (73.78%)	3.36 (29.29%)	3.15 (25.04%)	26.96 %	15.50 %	14.66 %
0.12 dB	4.78 (61.80%)	2.91 (21.00%)	2.74 (16.83%)	20.65 %	11.41 %	10.68 %
1.12 dB	4.07 (48.15%)	2.54 (14.33%)	2.38 (10.75%)	15.08 %	8.12 %	7.46 %
2.12 dB	3.36 (34.32%)	2.18 (8.52%)	2.09 (6.67%)	10.15 %	5.20 %	4.76 %

Table 4.3: Average Spectral Distortion and Symbol Error Rate for Uncoded LSP parameters over the AWGN channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

–	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-2.88 dB	7.67 (95.41%)	4.93 (57.75%)	4.71 (55.22%)	50.49 %	32.24 %	31.50 %
-1.88 dB	7.31 (93.45%)	4.57 (51.04%)	4.33 (47.36%)	46.20 %	28.35 %	27.53 %
-0.88 dB	6.89 (90.35%)	4.21 (45.19%)	3.92 (39.66%)	41.67 %	24.71 %	23.41 %
0.12 dB	6.50 (86.60%)	3.87 (38.21%)	3.58 (32.97%)	37.07 %	21.11 %	19.90 %
1.12 dB	6.10 (82.21%)	3.54 (31.83%)	3.29 (26.74%)	32.74 %	17.84 %	16.75 %
2.12 dB	5.67 (76.24%)	3.23 (25.54%)	3.01 (21.29%)	28.34 %	14.85 %	13.93 %
3.12 dB	5.24 (69.36%)	2.96 (20.66%)	2.77 (16.43%)	24.29 %	12.27 %	11.48 %
4.12 dB	4.81 (61.72%)	2.72 (16.43%)	2.54 (12.31%)	20.63 %	10.09 %	9.37 %
5.12 dB	4.40 (54.56%)	2.53 (13.55%)	2.36 (9.40%)	17.40 %	8.34 %	7.66 %
6.12 dB	4.01 (46.91%)	2.35 (10.75%)	2.20 (7.17%)	14.51 %	6.78 %	6.20 %

Table 4.4: Average Spectral Distortion and Symbol Error Rate for Uncoded LSP parameters over the Rayleigh Fading channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

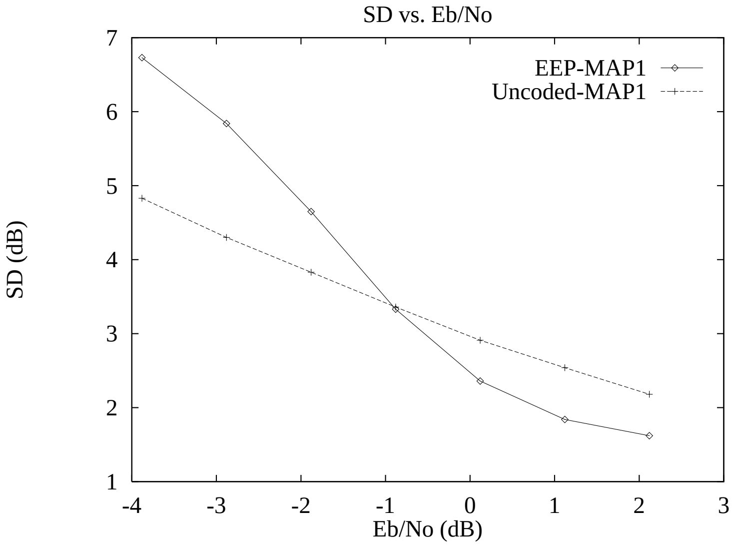

Figure 4.5: Spectral Distortion for Uncoded MAP1 vs EEP MAP1 Decoding Schemes of LSP parameters over the AWGN Channel.

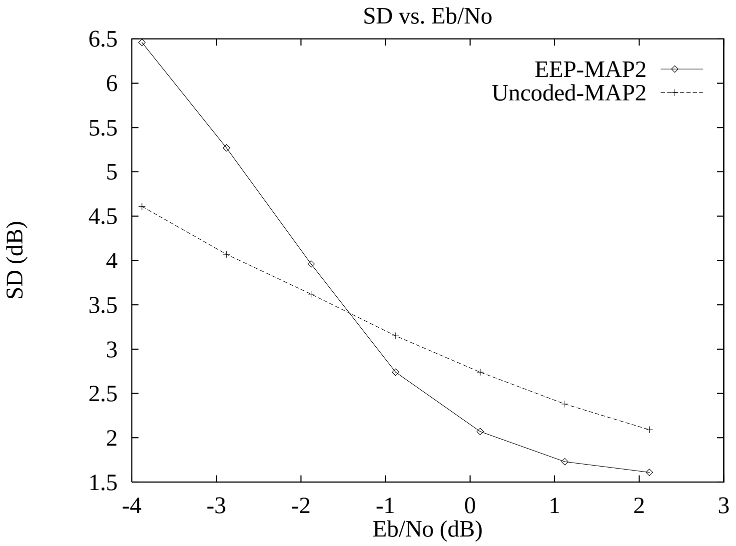

Figure 4.6: Spectral Distortion for Uncoded MAP2 vs EEP MAP2 Decoding Schemes of LSP parameters over the AWGN Channel.

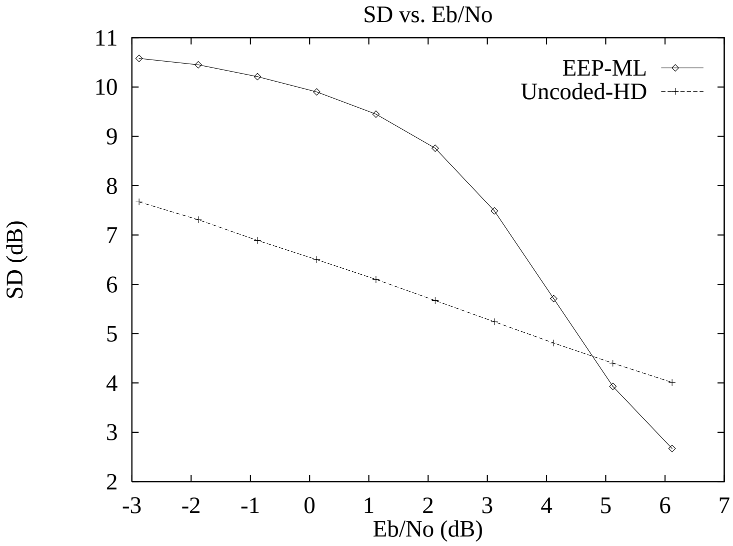

Figure 4.7: Spectral Distortion for Uncoded HD vs EEP ML Decoding Schemes of LSP parameters over the Rayleigh Channel.

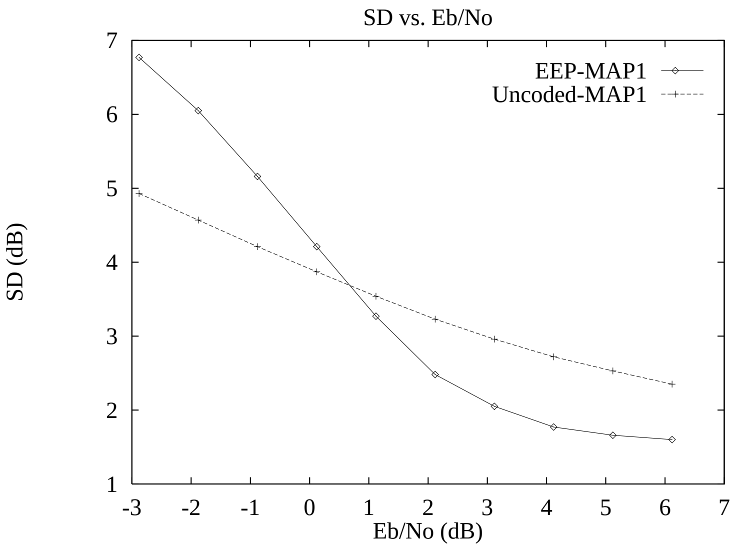

Figure 4.8: Spectral Distortion for Uncoded MAP1 vs EEP MAP1 Decoding Schemes of LSP parameters over the Rayleigh Channel.

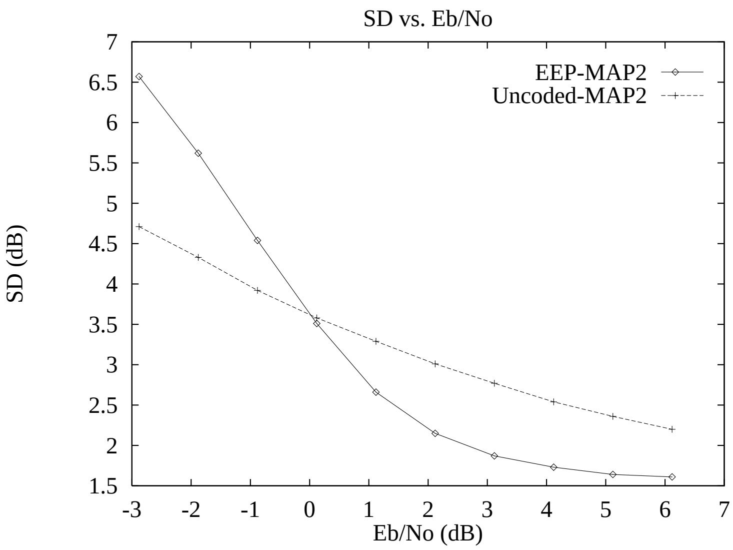

Figure 4.9: Spectral Distortion for Uncoded MAP2 vs EEP MAP2 Decoding Schemes of LSP parameters over the Rayleigh Channel.

As before, the LSB's of LSP 2 through 5 are not channel coded, and thus unordered LSP vectors must be reordered after decoding in order to yield a proper output.

One difference in the UEP system compared to the EEP is the path memory, p. In order, to keep the delay consistent between the UEP systems and the EEP system, the path memory is set accordingly to produce a delay of one frame (as before). The actual value of p will be specified for each RCPC family of codes, and will depend on the number of states of the mother code of the RCPC family.

As mentioned in the previous section, some of our UEP schemes will involve hybrid systems where some LSP parameters are sent without any channel coding. This will slightly increase decoding complexity of the specific UEP scheme, since parallel decoders must be employed.

Three sets of simulations are performed, one for each different mother code rate of the various RCPC families mentioned in Section 4.2. For each mother code rate 20 different systems were tested and the results of the best one are shown.

4.3.1 Rate 1/2 RCPC mother code simulations

The rate 1/2 mother codes that were used in our simulations were developed by Lee [20], [21]. Lee's RCPC systems were developed for periods between 2 and 7 and various different constraint lengths of mother codes. In our tests we only considered the families of codes based on a 32-state mother code, in order to keep decoder complexity the same as for the EEP simulations. We found that a period 5 family of RCPC codes outperformed the other families of codes (i.e., with different periods). The RCPC codes and protection schemes are summarized in Table 4.5 and Table 4.6, respectively.

Remark that the path memory used in this simulation is $p_{1/2} = 30$, which corresponds to a delay of one frame for a base rate 1/2 RCPC coder. This is actually

Mother Code Generator		53 75
Mother Code d_{free}		8
Memory Length		5
Period		5
Overall Rate		$\frac{34}{46}$
Rate	d_{free}	$A(\delta)$
5/9	6	11011 11111
5/8	6	11011 01111
5/7	5	11001 01111
5/6	3	10001 01111

Table 4.5: Summary of Rate 1/2 Family of RCPC Codes Used in the Simulations.

LSP	Code Rate
1	5/9
2-5	5/8
6-7	5/7
8-10	Uncoded

Table 4.6: LSP UEP Protection Scheme Using Rate 1/2 Family of RCPC Codes.

6 times the maximal memory element, where $m_{1/2} = 5$. This is *relatively* larger than the path memory for the EEP simulations (recall that $p_{EEP} = 10 = 5m_{EEP}$). Thus, the error due to path truncation in the decoder trellis will be reduced in this simulation. Also, the code's generator matrix is given in hexadecimal format, where $\mathbf{G}(D) = [53\ 75]$ corresponds to a generator matrix of,

$$\mathbf{G}(D) = [1 + D^2 + D^4 + D^5 \quad 1 + D^1 + D^2 + D^3 + D^5]. \tag{4.16}$$

Equation (3.55) was used to normalize the channel SNR's between this system and the EEP system. The results for the various performance criteria are summarized in Tables 4.11 to 4.14. Tables 4.15 to 4.20 show coding gains in E_b/N_0 for all three performance criteria. Figures 4.10 to 4.15 compare the performance of the base code rate 1/2 UEP system, with the EEP 3/4 and the uncoded system, for all three criteria and using the MAP2 decoding system.

Observations Regarding Results:

- It can be seen from Figures 4.10 to 4.15 that the UEP system performs better than the EEP Rate 3/4 system at lower E_b/N_0 values, yet at higher SNR's the EEP system seems to be slightly better. This is due to the use of uncoded LSP parameters in the UEP scheme (LSP8, LSP9 and LSP10), which have good performance at low SNR's, but not at high ones.

- At an average spectral distortion of 3.0 dB, a coding gain of 0.26 dB and 0.50 dB is seen for the AWGN and Rayleigh Fading channels, respectively. Similarly, at an average speech distortion of 6.0 dB, a coding gain of 0.81 dB and 1.04 dB is seen for the AWGN and Rayleigh Fading channels, respectively.

- Tables 4.17 and 4.20 show that gains for UEP - MAP2 vs EEP - MAP2 do not start until around 10% symbol error rate. This is because the UEP scheme attempts to reduce the number of symbol errors in the most important LSP's, and not necessarily the overall symbol error rate.

- The overall gains over the Rayleigh fading channel are greater than those of the AWGN channel. For example, at an average speech distortion of 5.5 dB, the total gain over the AWGN channel (Table 4.16) is 3.40 dB, compared to 5.91 dB for the Rayleigh channel (Table 4.19).

- Observe that coding gain is non-linear in nature. In other words, at high E_b/N_0 the distortion measures and the error rate are relatively horizontal for all the coding schemes, whereas they have a steep slope at lower SNR's. This will produce very large coding losses for a slight difference in decibels of distortion of percentage error rate at these high E_b/N_0 values.

4.3.2 Rate 1/3 RCPC mother code simulations

The rate 1/3 mother codes that were used in these simulations were developed by Hagenauer [14]. We used Hagenaur's rate 1/3 RCPC system, which is a 32-state system developed only for a period $p = 8$. The RCPC codes and chosen protection scheme are summarized in Table 4.7 and Table 4.8, respectively.

For conciseness the code rates that were used are summarized in Table 4.7, (for further details refer to [14]).

The path memory used in this simulation is also 30, ie. $p_{1/3} = 30$, which corresponds to a delay of one frame for a base rate 1/3 RCPC coder. As before, this corresponds to 6 times the maximal memory element, in other words $p_{1/3} = 6m_{1/3}$.

Once again, equation (3.55) was used to normalize the channel SNR's between this system and the EEP system. The results for the three performance criteria are summarized in Tables 4.21 to 4.24. Tables 4.25 to 4.30 show coding gains in E_b/N_0 for

Mother Code Generator		75 53 47
Mother Code d_{free}		13
Memory Length		5
Period		8
Overall Rate		$\frac{34}{66}$
Rate	d_{free}	$A(\delta)$
8/20	10	11111111 11111111 10101010
8/18	8	11111111 11111111 10001000
8/16	8	11111111 11111111 00000000

Table 4.7: Summary of Rate 1/3 Family of RCPC codes Used in the Simulations.

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

Table 4.8: LSP UEP Protection Scheme Using Rate 1/3 Family of RCPC Codes.

all the previously mentioned performance criteria. Figures 4.16 to 4.21 compare the performance of the base code rate 1/3 UEP system, with the EEP 3/4 and uncoded system, again using the MAP2 decoding system. Note that in these Figures, an EEP rate 1/2 system[‡] is also depicted. Although, all the different system's are normalized for the sake of comparison, this is done to show a valid comparison to a system with a *similar* rate.

Observations Regarding Results:

- Higher gains are achieved with this family of codes relative to the previous rate 1/2 family of RCPC codes. More specifically, at spectral distortion equal to 3.0 dB, a coding gain for UEP vs. EEP (MAP2) of 0.81 dB and 1.54 dB is seen for the AWGN and Rayleigh fading channels, respectively. These are 0.55 dB and 1.04 dB greater than gains at the same spectral distortion for the rate 1/2 UEP system. This is due to the more powerful mother code, which allows lower rate (higher d_{free}) codes to be used on the more important parameters.

- At an average speech distortion of 6.0 dB, a coding gain of 1.05 dB and 1.88 dB is seen for the AWGN and Rayleigh Fading channels, respectively.

- We see significant coding gains for the UEP system over the EEP-MAP2 system in P_s at error rates as low as 5%. For example, for the Rayleigh fading channel, there is a 1.07 dB gain at $P_s = 5\%$. The fact that gains are starting at lower values of symbol error rate is due to the Markov structure of our MAP model; since we are giving better protection to the first LSP this will result in a more effective implementation of MAP decoding.

- Figures 4.16 to 4.21 demonstrate that the gains for the UEP system are not as pronounced when compared to a similar rate EEP system (EEP rate 1/2

‡Actually this is the unpunctured mother code of the rate 1/2 family of RCPC codes.

in this case). However, the results are still important; they provide significant improvements in terms of the three performance criteria.

4.3.3 Rate 1/4 RCPC mother code simulations

Hagenauer's rate 1/4 RCPC system [14], which is a 16-state system developed only for a period $p = 8$, was next implemented§. The RCPC codes and chosen protection scheme are summarized in Table 4.9 and Table 4.10, respectively.

The path memory used in this simulation is also 30, ie. $p_{1/4} = 30$, which corresponds to a delay of one frame for a base rate 1/4 RCPC coder. However, due to a shorter memory length, this corresponds to 7.5 times the maximal memory element, in other words $p_{1/4} = 7.5m_{1/4}$.

The results are summarized in Tables 4.31 to 4.34, after Equation (3.55) was used to normalize the systems. Tables 4.35 to 4.40 show coding gains in E_b/N_0 for the UEP system versus the rate 3/4 EEP system with and without MAP2 decoding. Figures 4.22 to 4.27 compare the performance of the base code rate 1/4 UEP system, with the EEP 3/4 and uncoded system, again using the MAP2 decoding system.

- Even higher gains are achieved with this RCPC family of codes than the previous ones. More specifically, at spectral distortion equal to 3.0 dB, coding gains for UEP vs. EEP (MAP2) of 0.95 dB and 2.11 dB are observed for the AWGN and Rayleigh fading channels, respectively.

- This system provides the highest overall gains for UEP versus EEP - ML, with over 8 dB gains (Table 4.38) achieved under the average spectral distortion criterion over the Rayleigh channel.

§A 32-state mother code rate 1/4 RCPC family of codes has not been developed.

Mother Code Generator		23 35 27 33
Mother Code d_{free}		15
Memory Length		4
Period		8
Overall Rate		$\frac{34}{105}$
Rate	d_{free}	$A(\delta)$
8/30	13	11111111 11111111 11111111 11101110
8/28	13	11111111 11111111 11111111 10101010
8/24	11	11111111 11111111 11111111 00000000

Table 4.9: Summary of Rate 1/4 Family of RCPC Codes Used in the Simulations.

LSP	Code Rate
1-4	8/30
5	8/28
6-10	8/24

Table 4.10: LSP UEP Protection Scheme Using Rate 1/4 Family of RCPC Codes.

4.3.4 Conclusions

The unequal error protection schemes perform better than the equal error protection schemes at low values of SNR. For the Rayleigh channel, coding gains of up to 2.2 dB for the same average spectral distortion and up to 2.5 dB for the same average speech distortion are realized with rate-1/4 unequal error protection over rate-3/4 equal error protection, with both systems using MAP channel decoding.

At mid-range values of SNR we note that the two types of coding schemes perform similarly, with the equal error protection showing better performance at high E_b/N_0.

We also observe a smoother degradation in performance as the channel becomes noisier for the unequal error protection schemes. All these observations demonstrate that unequal error protection is advantageous in very noisy channel conditions.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-3.88 dB	10.60	7.63	7.28
-2.88 dB	10.17	6.85	6.45
-1.88 dB	9.43	6.10	5.76
-0.88 dB	8.23	5.56	5.35
0.12 dB	6.74	5.20	5.13
1.12 dB	5.72	5.05	5.02
2.12 dB	5.21	4.96	4.95

Table 4.11: Average Speech Distortion for Rate-1/2 RCPC Unequal Error Protection of LSP Parameters over the AWGN Channel.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-2.88 dB	10.62	7.64	7.27
-1.88 dB	10.37	7.08	6.64
-0.88 dB	9.99	6.45	6.04
0.12 dB	9.37	5.91	5.60
1.12 dB	8.46	5.51	5.34
2.12 dB	7.25	5.28	5.17
3.12 dB	6.29	5.15	5.08
4.12 dB	5.67	5.06	5.02
5.12 dB	5.34	5.01	4.98
6.12 dB	5.19	4.97	4.94

Table 4.12: Average Speech Distortion for Rate-1/2 RCPC Unequal Error Protection of LSP Parameters over the Rayleigh Fading Channel.

–	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.88 dB	9.82 (98.07%)	5.90 (61.64%)	5.28 (54.19%)	71.99 %	36.39 %	32.81 %
-2.88 dB	9.35 (96.10%)	4.89 (47.95%)	4.26 (40.00%)	66.53 %	27.15 %	23.50 %
-1.88 dB	8.50 (89.66%)	3.84 (33.35%)	3.33 (25.60%)	57.20 %	18.34 %	14.92 %
-0.88 dB	7.03 (76.98%)	3.04 (21.38%)	2.70 (15.86%)	42.65 %	11.73 %	9.59 %
0.12 dB	5.03 (53.47%)	2.45 (12.33%)	2.30 (9.26%)	25.18 %	7.11 %	6.28 %
1.12 dB	3.44 (31.98%)	2.15 (7.55%)	2.08 (6.14%)	12.40 %	4.72 %	4.35 %
2.12 dB	2.55 (17.47%)	1.94 (4.52%)	1.89 (3.62%)	5.97 %	3.16 %	2.91 %

Table 4.13: Average Spectral Distortion and Symbol Error Rate for Rate-1/2 RCPC Unequal Error Protection of LSP Parameters over the AWGN Channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

–	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-2.88 dB	9.92 (98.33%)	5.85 (60.06%)	5.19 (53.04%)	73.53 %	36.54 %	32.68 %
-1.88 dB	9.61 (97.46%)	5.11 (50.32%)	4.42 (41.69%)	69.96 %	29.50 %	25.05 %
-0.88 dB	9.17 (94.40%)	4.26 (38.42%)	3.68 (30.82%)	64.37 %	22.01 %	18.02 %
0.12 dB	8.46 (89.51%)	3.52 (27.92%)	3.04 (20.41%)	56.60 %	15.80 %	12.69 %
1.12 dB	7.30 (79.56%)	2.95 (19.34%)	2.65 (14.83%)	44.96 %	10.98 %	9.29 %
2.12 dB	5.75 (63.46%)	2.56 (13.25%)	2.35 (9.89%)	31.22 %	7.93 %	6.80 %
3.12 dB	4.37 (45.98%)	2.33 (9.99%)	2.18 (6.94%)	19.17 %	6.18 %	5.40 %
4.12 dB	3.40 (32.04%)	2.17 (7.83%)	2.06 (5.34%)	11.45 %	5.00 %	4.40 %
5.12 dB	2.86 (23.16%)	2.06 (6.25%)	1.97 (4.25%)	7.64 %	4.10 %	3.62 %
6.12 dB	2.57 (18.01%)	1.96 (5.01%)	1.88 (3.28%)	5.80 %	3.37 %	2.92 %

Table 4.14: Average Spectral Distortion and Symbol Error Rate for Rate-1/2 RCPC Unequal Error Protection of LSP Parameters over the Rayleigh Fading Channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

Coding Gains	Avg. SD (dB)			
	2.0	2.5	3.0	3.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+2.16	+2.81	+3.18	+3.39
MAP2 – UEP 1/2 vs. MAP2 – EEP 3/4	-1.22	-0.14	+0.26	+0.56
Total Gain MAP2 – UEP 1/2 vs. ML – EEP 3/4	+0.94	+2.66	+3.44	+3.95

Table 4.15: Coding Gains for Base Rate-1/2 Unequal Error Protection over the AWGN channel for the Same Average Spectral Distortion.

Coding Gains	Avg. Speech Dist. (dB)			
	5.0	5.5	6.0	6.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+1.84	+2.97	+3.27	+3.51
MAP2 – UEP 1/2 vs. MAP2 – EEP 3/4	-0.79	+0.44	+0.81	+0.97
Total Gain MAP2 – UEP 1/2 vs. ML – EEP 3/4	+1.06	+3.40	+4.08	+4.49

Table 4.16: Coding Gains for Base Rate-1/2 Unequal Error Protection over the AWGN channel for the Same Average Speech Distortion.

Coding Gains	P_s (%)				
	1 %	5%	10%	15%	20%
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+1.34	+2.47	+3.07	+3.33	+3.59
MAP2 – UEP 1/2 vs. MAP2 – EEP 3/4	-2.24	-0.87	+0.06	+0.54	+0.66
Total Gain MAP2 – UEP 1/2 vs. ML – EEP 3/4	-0.90	+1.60	+3.14	+3.87	+4.25

Table 4.17: Coding Gains for Base Rate-1/2 Unequal Error Protection over the AWGN channel for the Same Symbol Error rate.

Coding Gains	Avg. SD (dB)			
	2.0	2.5	3.0	3.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+4.79	+5.62	+5.94	+6.13
MAP2 – UEP 1/2 vs. MAP2 – EEP 3/4	-2.13	-0.19	+0.50	+0.73
Total Gain MAP2 – UEP 1/2 vs. ML – EEP 3/4	+2.66	+5.43	+6.43	+6.86

Table 4.18: Coding Gains for Base Rate-1/2 Unequal Error Protection over the Rayleigh Fading channel for the Same Average Spectral Distortion.

Coding Gains	Avg. Speech Dist. (dB)			
	5.0	5.5	6.0	6.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+3.99	+5.36	+5.64	+5.79
MAP2 – UEP 1/2 vs. MAP2 – EEP 3/4	-1.68	+0.55	+1.04	+1.23
Total Gain MAP2 – UEP 1/2 vs. ML – EEP 3/4	+2.31	+5.91	+6.68	+7.02

Table 4.19: Coding Gains for Base Rate-1/2 Unequal Error Protection over the Rayleigh Fading channel for the Same Average Speech Distortion.

Coding Gains	P_s (%)				
	1 %	5%	10%	15%	20%
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+2.78	+5.02	+5.65	+5.91	+6.09
MAP2 – UEP 1/2 vs. MAP2 – EEP 3/4	-4.80	-1.55	+0.06	+0.65	+0.94
Total Gain MAP2 – UEP 1/2 vs. ML – EEP 3/4	-2.01	+3.47	+5.70	+6.56	+7.03

Table 4.20: Coding Gains for Base Rate-1/2 Unequal Error Protection over the Rayleigh Fading channel for the Same Symbol Error rate.

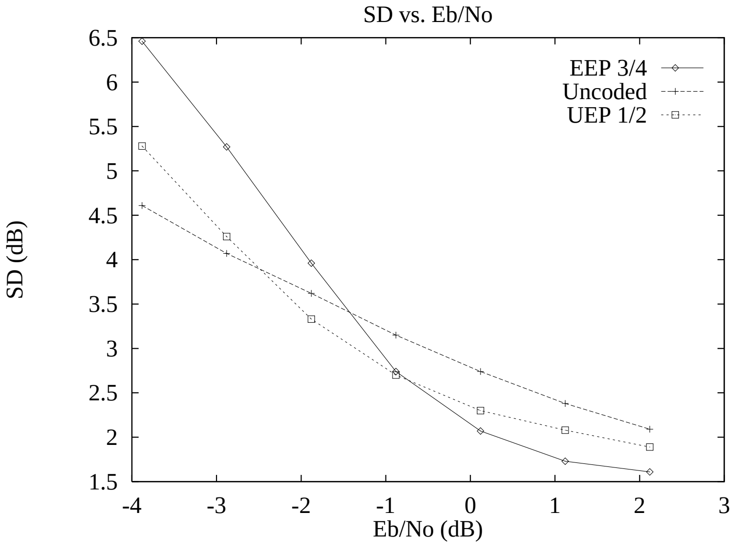

Figure 4.10: Spectral Distortion for Rate-1/2 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

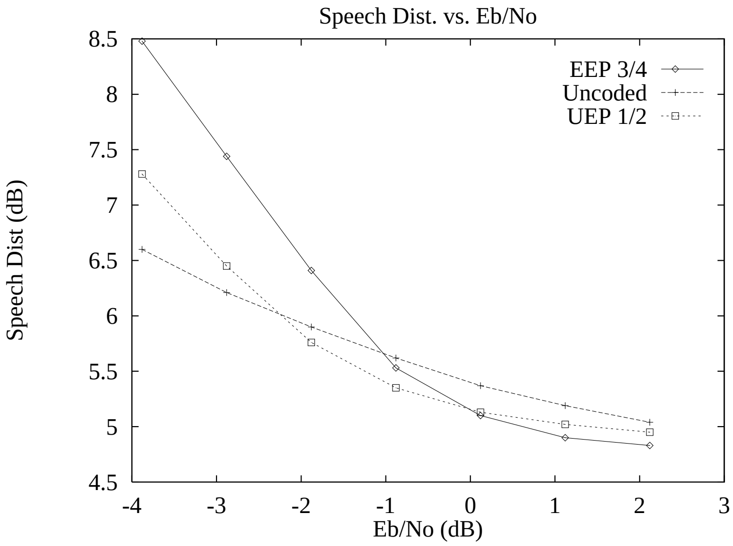

Figure 4.11: Average Speech Distortion for Rate-1/2 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

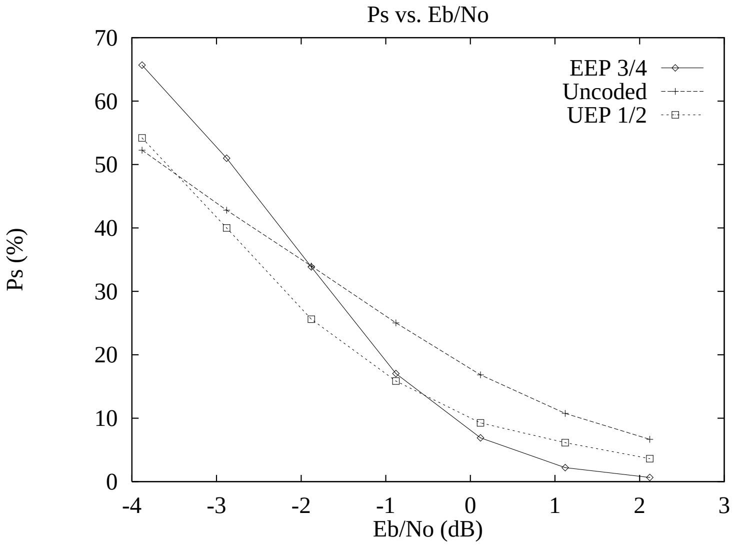

Figure 4.12: Symbol Error Rate for Rate-1/2 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

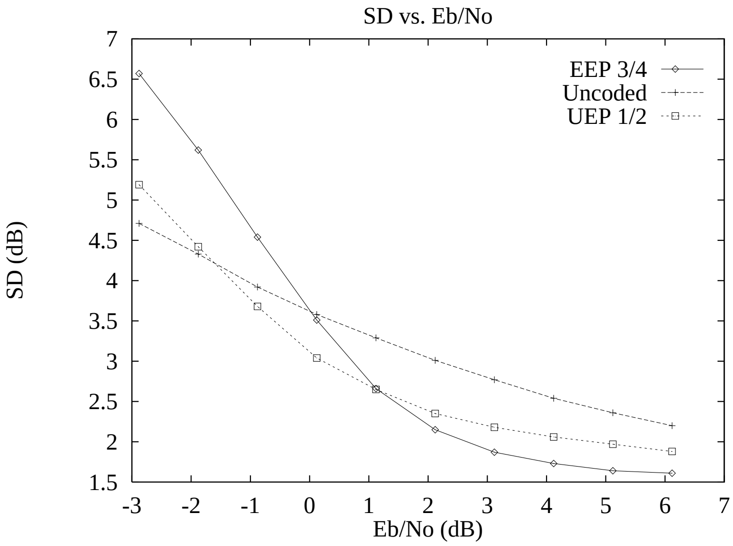

Figure 4.13: Spectral Distortion for Rate-1/2 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

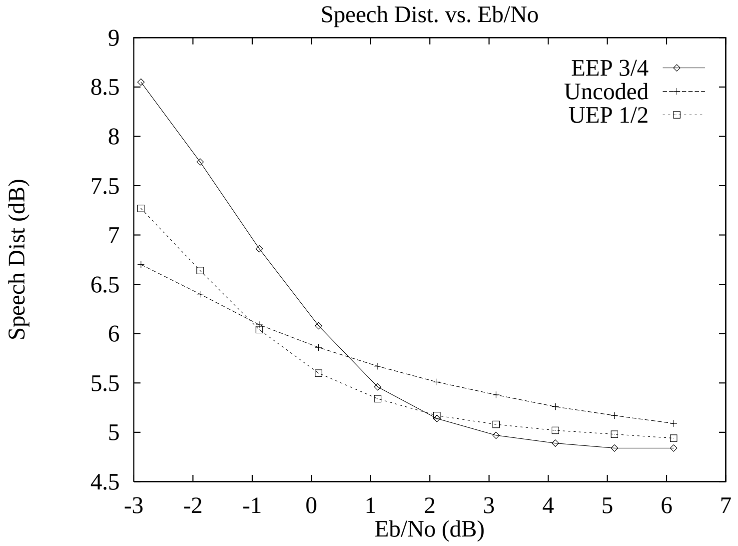

Figure 4.14: Average Speech Distortion for Rate-1/2 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

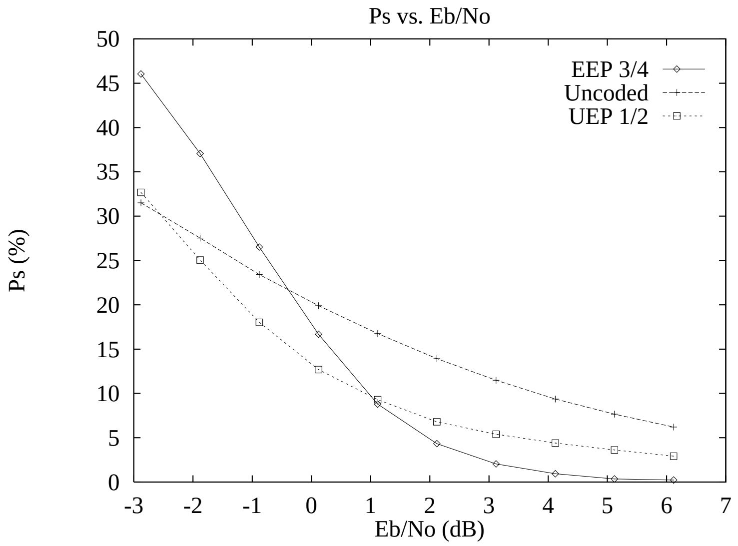

Figure 4.15: Symbol Error Rate for Rate-1/2 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-3.88 dB	10.50	7.47	7.33
-2.88 dB	9.81	6.68	6.31
-1.88 dB	8.73	5.89	5.56
-0.88 dB	7.25	5.34	5.20
0.12 dB	5.90	5.05	5.01
1.12 dB	5.17	4.94	4.94
2.12 dB	4.97	4.90	4.90

Table 4.21: Average Speech Distortion for Rate-1/3 RCPC Unequal Error Protection of LSP Parameters over the AWGN Channel.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-2.88 dB	10.58	7.30	6.89
-1.88 dB	9.98	6.59	6.13
-0.88 dB	9.06	5.94	5.61
0.12 dB	7.78	5.43	5.26
1.12 dB	6.48	5.13	5.08
2.12 dB	5.56	5.00	4.98
3.12 dB	5.16	4.94	4.93
4.12 dB	5.01	4.91	4.91
5.12 dB	4.95	4.89	4.89
6.12 dB	4.92	4.88	4.87

Table 4.22: Average Speech Distortion for Rate-1/3 RCPC Unequal Error Protection of LSP Parameters over the Rayleigh Fading Channel.

–	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.88 dB	9.75 (96.61%)	5.69 (57.49%)	5.26 (52.39%)	73.41 %	35.03 %	31.96 %
-2.88 dB	9.02 (91.35%)	4.58 (42.05%)	3.98 (34.80%)	65.07 %	24.47 %	20.67 %
-1.88 dB	7.67 (79.08%)	3.45 (26.26%)	2.98 (19.23%)	50.76 %	14.66 %	11.50 %
-0.88 dB	5.63 (55.97%)	2.61 (13.07%)	2.39 (9.26%)	31.69 %	7.63 %	6.35 %
0.12 dB	3.61 (29.44%)	2.15 (5.76%)	2.06 (4.21%)	14.44 %	4.12 %	3.59 %
1.12 dB	2.45 (12.38%)	1.94 (2.95%)	1.92 (2.59%)	5.57 %	2.64 %	2.59 %
2.12 dB	2.05 (6.19%)	1.83 (2.00%)	1.81 (1.54%)	2.94 %	1.98 %	1.89 %

Table 4.23: Average Spectral Distortion and Symbol Error Rate for Rate-1/3 RCPC Unequal Error Protection of LSP Parameters over the AWGN Channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

–	SD (dB)			P_s(%)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-2.88 dB	9.86 (96.33%)	5.44 (53.78%)	4.79 (46.76%)	73.73 %	32.10 %	28.17 %
-1.88 dB	9.19 (92.13%)	4.46 (40.82%)	3.79 (31.85%)	66.38 %	23.05 %	18.84 %
-0.88 dB	8.12 (83.08%)	3.52 (26.78%)	3.03 (19.46%)	54.61 %	14.83 %	11.64 %
0.12 dB	6.44 (65.29%)	2.78 (15.57%)	2.52 (11.15%)	38.26 %	8.69 %	6.97 %
1.12 dB	4.60 (43.49%)	2.32 (8.35%)	2.22 (6.54%)	21.97 %	5.12 %	4.44 %
2.12 dB	3.15 (23.65%)	2.08 (4.63%)	2.03 (3.70%)	10.31 %	3.29 %	3.04 %
3.12 dB	2.48 (13.19%)	1.96 (3.22%)	1.93 (2.52%)	5.38 %	2.54 %	2.37 %
4.12 dB	2.19 (8.75%)	1.90 (2.69%)	1.87 (2.10%)	3.49 %	2.18 %	2.01 %
5.12 dB	2.06 (6.84%)	1.84 (2.29%)	1.82 (1.85%)	2.78 %	1.82 %	1.73 %
6.12 dB	1.97 (5.45%)	1.80 (2.15%)	1.76 (1.33%)	2.34 %	1.59 %	1.39 %

Table 4.24: Average Spectral Distortion and Symbol Error Rate for Rate-1/3 RCPC Unequal Error Protection of LSP Parameters over the Rayleigh Fading Channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

Coding Gains	Avg. SD (dB)			
	2.0	2.5	3.0	3.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+2.16	+2.81	+3.18	+3.39
MAP2 – UEP 1/3 vs. MAP2 – EEP 3/4	-0.22	+0.54	+0.81	+0.90
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 3/4	+1.93	+3.35	+3.98	+4.28

Table 4.25: Coding Gains for Base Rate-1/3 Unequal Error Protection over the AWGN channel for the Same Average Spectral Distortion.

Coding Gains	Avg. Speech Dist. (dB)			
	5.0	5.5	6.0	6.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+1.84	+2.97	+3.27	+3.51
MAP2 – UEP 1/3 vs. MAP2 – EEP 3/4	+0.36	+0.90	+1.05	+1.10
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 3/4	+2.20	+3.87	+4.32	+4.61

Table 4.26: Coding Gains for Base Rate-1/3 Unequal Error Protection over the AWGN channel for the Same Average Speech Distortion.

Coding Gains	P_s (%)				
	1 %	5%	10%	15%	20%
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+1.34	+2.47	+3.07	+3.33	+3.59
MAP2 – UEP 1/3 vs. MAP2 – EEP 3/4	-2.18	+0.30	+0.69	+0.91	+1.00
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 3/4	-0.85	+2.77	+3.77	+4.24	+4.58

Table 4.27: Coding Gains for Base Rate-1/3 Unequal Error Protection over the AWGN channel for the Same Symbol Error rate.

Coding Gains	Avg. SD (dB)			
	2.0	2.5	3.0	3.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+4.79	+5.62	+5.94	+6.13
MAP2 – UEP 1/3 vs. MAP2 – EEP 3/4	+0.24	+1.25	+1.54	+1.63
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 3/4	+5.02	+6.86	+7.48	+7.76

Table 4.28: Coding Gains for Base Rate-1/3 Unequal Error Protection over the Rayleigh Fading channel for the Same Average Spectral Distortion.

Coding Gains	Avg. Speech Dist. (dB)			
	5.0	5.5	6.0	6.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+3.99	+5.36	+5.64	+5.79
MAP2 – UEP 1/3 vs. MAP2 – EEP 3/4	+1.02	+1.62	+1.88	+1.95
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 3/4	+5.01	+6.98	+7.52	+7.74

Table 4.29: Coding Gains for Base Rate-1/3 Unequal Error Protection over the Rayleigh Fading channel for the Same Average Speech Distortion.

Coding Gains	P_s (%)				
	1 %	5%	10%	15%	20%
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+2.78	+5.02	+5.65	+5.91	+6.09
MAP2 – UEP 1/3 vs. MAP2 – EEP 3/4	-3.20	+1.07	+1.50	+1.68	+1.79
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 3/4	-0.42	+6.09	+7.14	+7.59	+7.88

Table 4.30: Coding Gains for Base Rate-1/3 Unequal Error Protection over the Rayleigh Fading channel for the Same Symbol Error rate.

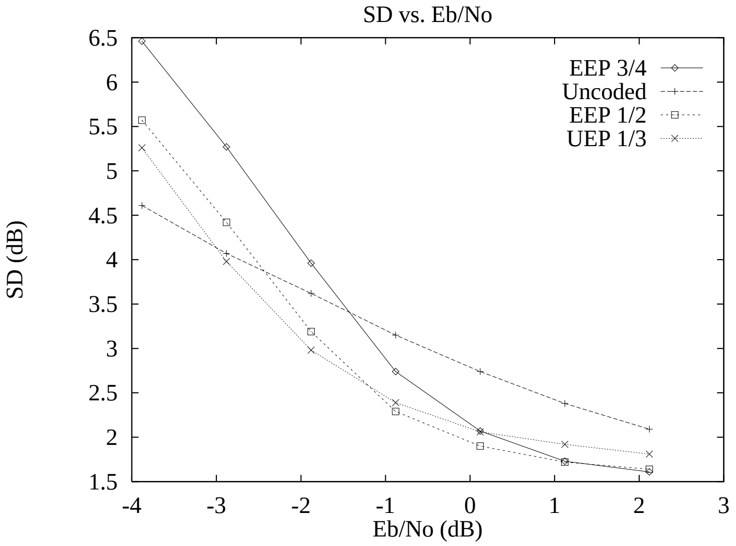

Figure 4.16: Spectral Distortion for Rate-1/3 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

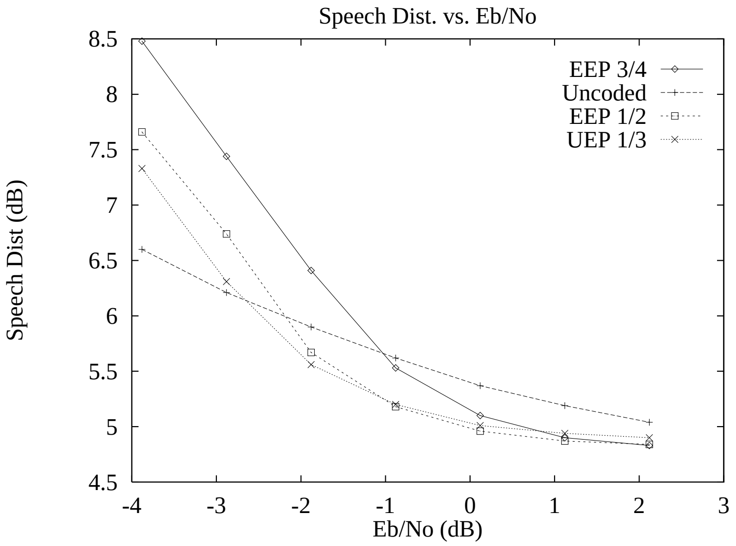

Figure 4.17: Average Speech Distortion for Rate-1/3 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

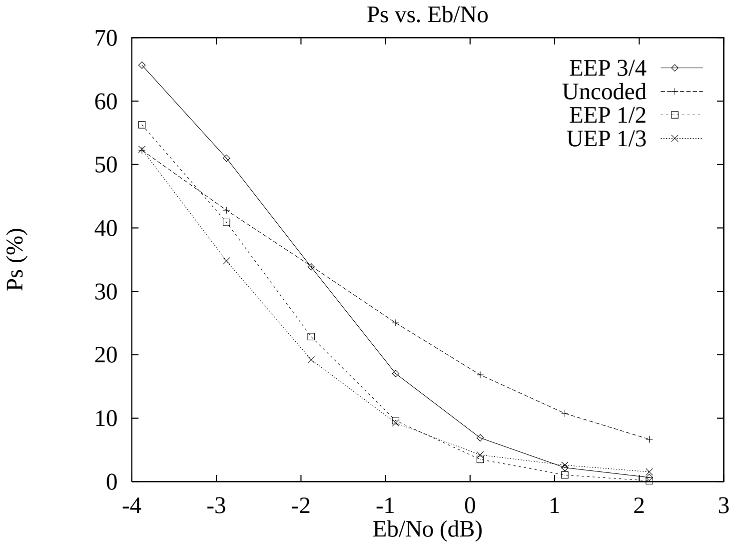

Figure 4.18: Symbol Error Rate for Rate-1/3 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

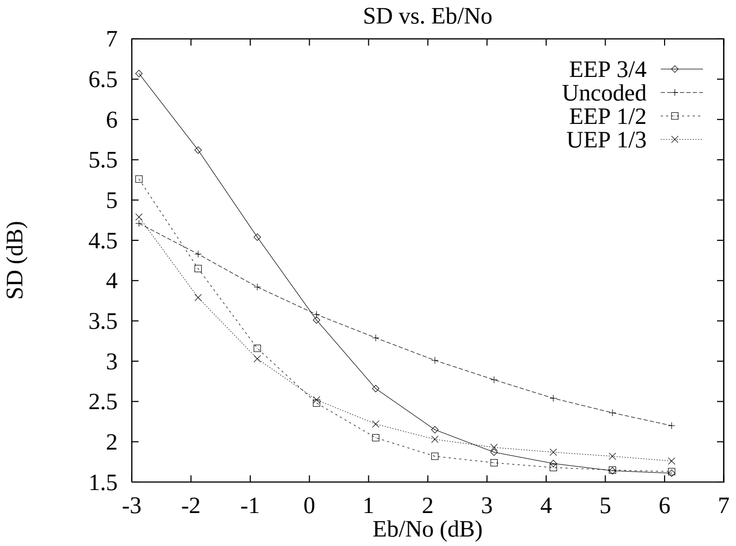

Figure 4.19: Spectral Distortion for Rate-1/3 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

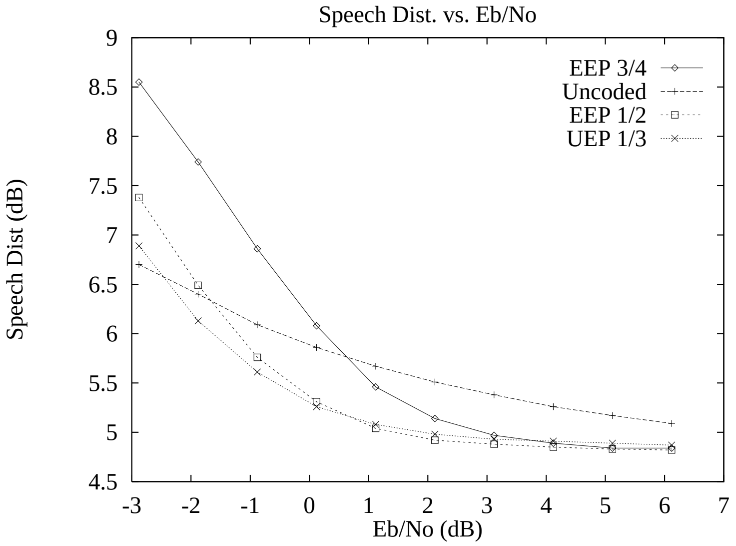

Figure 4.20: Average Speech Distortion for Rate-1/3 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

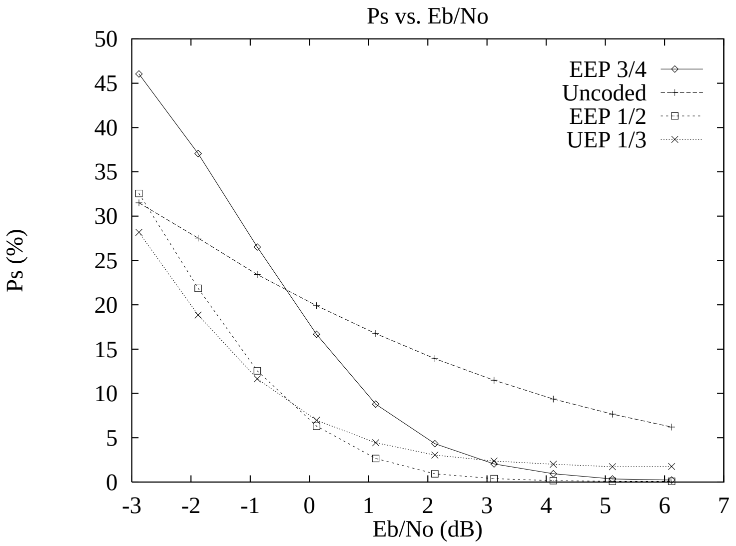

Figure 4.21: Symbol Error Rate for Rate-1/3 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-3.88 dB	9.81	7.15	6.85
-2.88 dB	8.89	6.39	6.08
-1.88 dB	7.66	5.77	5.50
-0.88 dB	6.41	5.28	5.14
0.12 dB	5.51	5.02	4.96
1.12 dB	5.04	4.92	4.90
2.12 dB	4.89	4.87	4.87

Table 4.31: Average Speech Distortion for Rate 1/4 RCPC Unequal Error Protection of LSP Parameters over the AWGN Channel.

–	Speech Distortion (dB)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2
-2.88 dB	9.51	6.79	6.48
-1.88 dB	8.59	6.17	5.85
-0.88 dB	7.41	5.61	5.39
0.12 dB	6.30	5.25	5.10
1.12 dB	5.50	5.04	4.97
2.12 dB	5.08	4.92	4.91
3.12 dB	4.93	4.88	4.88
4.12 dB	4.88	4.86	4.86
5.12 dB	4.85	4.85	4.85
6.12 dB	4.84	4.84	4.84

Table 4.32: Average Speech Distortion for Rate 1/4 RCPC Unequal Error Protection of LSP Parameters over the Rayleigh Fading Channel.

–	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.88 dB	9.04 (92.57%)	5.28 (54.27%)	4.75 (47.87%)	64.02 %	29.70 %	27.05 %
-2.88 dB	7.88 (82.53%)	4.24 (39.92%)	3.70 (31.73%)	51.81 %	20.39 %	17.02 %
-1.88 dB	6.22 (65.20%)	3.30 (25.35%)	2.86 (18.56%)	35.91 %	12.05 %	9.23 %
-0.88 dB	4.39 (41.74%)	2.49 (12.46%)	2.26 (8.50%)	19.66 %	5.38 %	3.99 %
0.12 dB	2.87 (19.14%)	2.02 (4.67%)	1.93 (3.07%)	7.90 %	1.91 %	1.40 %
1.12 dB	2.07 (5.98%)	1.82 (1.33%)	1.80 (1.07%)	2.28 %	0.57 %	0.46 %
2.12 dB	1.78 (1.43%)	1.72 (0.38%)	1.72 (0.44%)	0.43 %	0.13 %	0.14 %

Table 4.33: Average Spectral Distortion and Symbol Error Rate for Rate 1/4 RCPC Unequal Error Protection of LSP Parameters over the AWGN Channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

–	SD (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-2.88 dB	8.65 (89.39%)	4.79 (47.21%)	4.23 (40.30%)	59.89 %	25.33 %	21.93 %
-1.88 dB	7.46 (77.76%)	3.85 (33.37%)	3.33 (25.17%)	47.46 %	16.68 %	13.07 %
-0.88 dB	5.82 (60.16%)	3.00 (20.35%)	2.65 (13.91%)	32.37 %	9.42 %	7.11 %
0.12 dB	4.16 (38.25%)	2.41 (10.67%)	2.19 (6.73%)	17.67 %	4.62 %	3.17 %
1.12 dB	2.90 (19.12%)	2.04 (4.88%)	1.94 (2.84%)	7.84 %	1.85 %	1.25 %
2.12 dB	2.16 (7.58%)	1.84 (1.45%)	1.82 (1.14%)	2.63 %	0.56 %	0.48 %
3.12 dB	1.85 (2.23%)	1.76 (0.50%)	1.76 (0.53%)	0.69 %	0.17 %	0.19 %
4.12 dB	1.74 (0.72%)	1.71 (0.23%)	1.72 (0.29%)	0.16 %	0.04 %	0.07 %
5.12 dB	1.69 (0.23%)	1.68 (0.17%)	1.69 (0.27%)	0.03 %	0.01 %	0.05 %
6.12 dB	1.66 (0.15%)	1.66 (0.15%)	1.67 (0.23%)	0.01 %	0.01 %	0.05 %

Table 4.34: Average Spectral Distortion and Symbol Error Rate for Rate 1/4 RCPC Unequal Error Protection of LSP Parameters over the Rayleigh Fading Channel (Values in Parenthesis are Percentages of Outliers > 4 dB).

Coding Gains	Avg. SD (dB)			
	2.0	2.5	3.0	3.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+2.16	+2.81	+3.18	+3.39
MAP2 – UEP 1/4 vs. MAP2 – EEP 3/4	+0.42	+0.76	+0.95	+1.14
Total Gain MAP2 – UEP 1/4 vs. ML – EEP 3/4	+2.57	+3.56	+4.13	+4.53

Table 4.35: Coding Gains for Base Rate 1/4 Unequal Error Protection over the AWGN Channel for the Same Average Spectral Distortion.

Coding Gains	Avg. Speech Dist. (dB)			
	5.0	5.5	6.0	6.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+1.84	+2.97	+3.27	+3.51
MAP2 – UEP 1/4 vs. MAP2 – EEP 3/4	+0.72	+1.07	+1.33	+1.46
Total Gain MAP2 – UEP 1/4 vs. ML – EEP 3/4	+2.56	+4.04	+4.59	+4.97

Table 4.36: Coding Gains for Base Rate 1/4 Unequal Error Protection over the AWGN Channel for the Same Average Speech Distortion.

Coding Gains	P_s (%)				
	1 %	5%	10%	15%	20%
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+1.34	+2.47	+3.07	+3.33	+3.59
MAP2 – UEP 1/4 vs. MAP2 – EEP 3/4	+0.66	+0.98	+1.08	+1.27	+1.37
Total Gain MAP2 – UEP 1/4 vs. ML – EEP 3/4	+2.00	+3.46	+4.16	+4.60	+4.95

Table 4.37: Coding Gains for Base Rate 1/4 Unequal Error Protection over the AWGN Channel for the Same Symbol Error rate.

Coding Gains	Avg. SD (dB)			
	2.0	2.5	3.0	3.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+4.79	+5.62	+5.94	+6.13
MAP2 – UEP 1/4 vs. MAP2 – EEP 3/4	+1.78	+1.99	+2.11	+2.20
Total Gain MAP2 – UEP 1/4 vs. ML – EEP 3/4	+6.56	+7.60	+8.05	+8.33

Table 4.38: Coding Gains for Base Rate 1/4 Unequal Error Protection over the Rayleigh Fading Channel for the Same Average Spectral Distortion.

Coding Gains	Avg. Speech Dist. (dB)			
	5.0	5.5	6.0	6.5
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+3.99	+5.36	+5.64	+5.79
MAP2 – UEP 1/4 vs. MAP2 – EEP 3/4	+2.05	+2.17	+2.37	+2.49
Total Gain MAP2 – UEP 1/4 vs. ML – EEP 3/4	+6.04	+7.53	+8.01	+8.28

Table 4.39: Coding Gains for Base Rate 1/4 Unequal Error Protection over the Rayleigh Fading Channel for the Same Average Speech Distortion.

Coding Gains	P_s (%)				
	1 %	5%	10%	15%	20%
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+2.78	+5.02	+5.65	+5.91	+6.09
MAP2 – UEP 1/4 vs. MAP2 – EEP 3/4	+2.62	+2.31	+2.33	+2.43	+2.44
Total Gain MAP2 – UEP 1/4 vs. ML – EEP 3/4	+5.41	+7.33	+7.98	+8.34	+8.53

Table 4.40: Coding Gains for Base Rate 1/4 Unequal Error Protection over the Rayleigh Fading Channel for the Same Symbol Error rate.

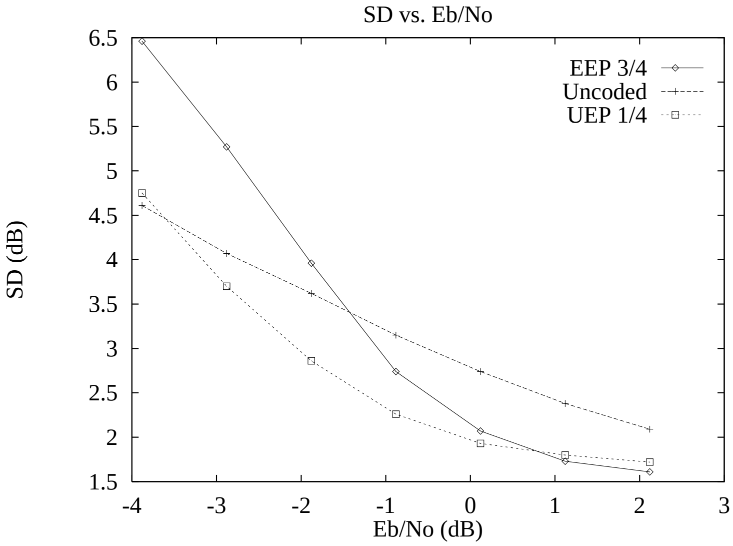

Figure 4.22: Spectral Distortion for Rate 1/4 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

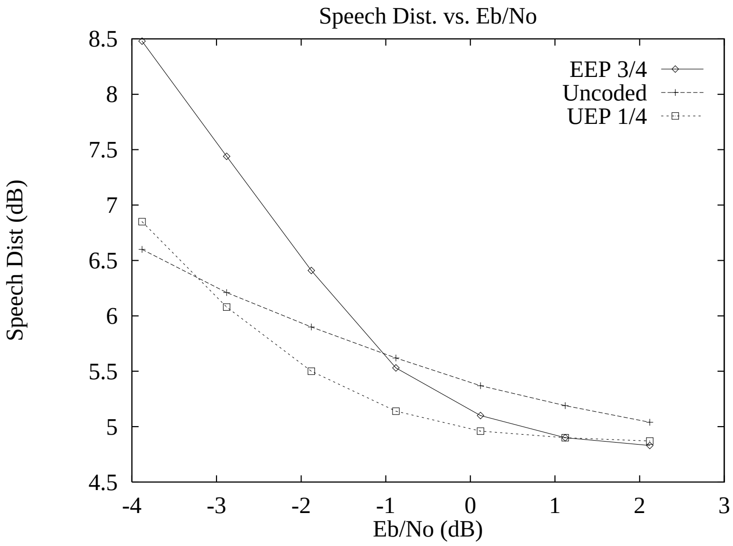

Figure 4.23: Average Speech Distortion for Rate 1/4 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

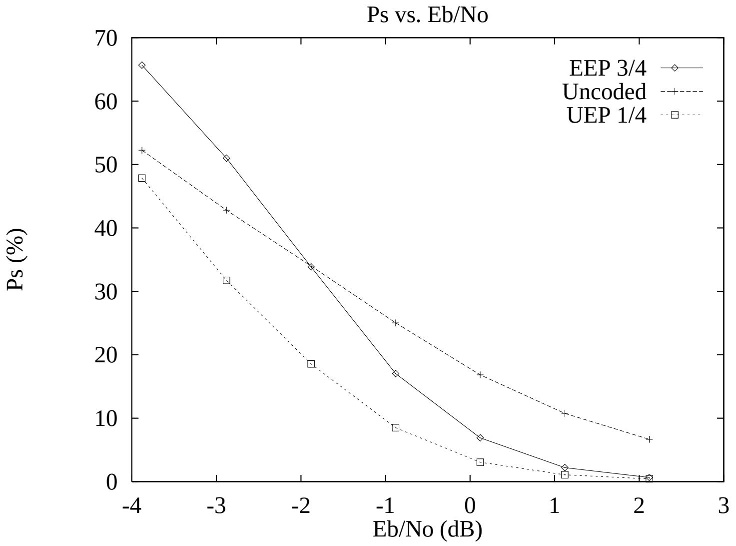

Figure 4.24: Symbol Error Rate for Rate 1/4 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the AWGN Channel.

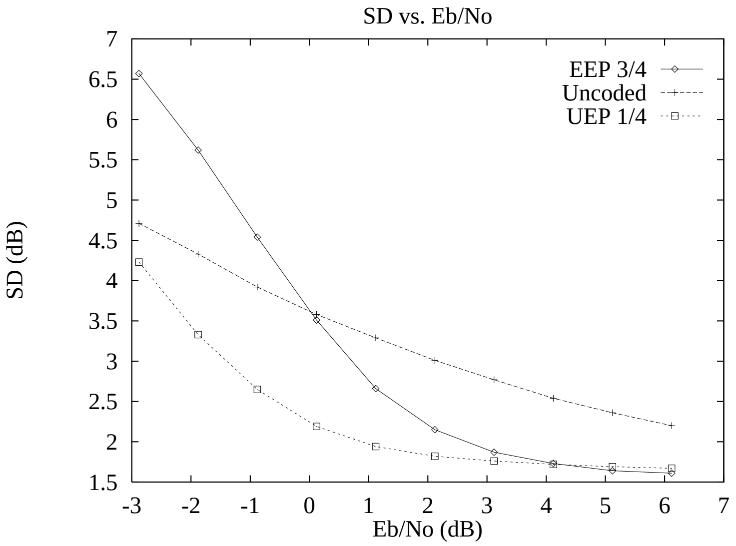

Figure 4.25: Spectral Distortion for Rate 1/4 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

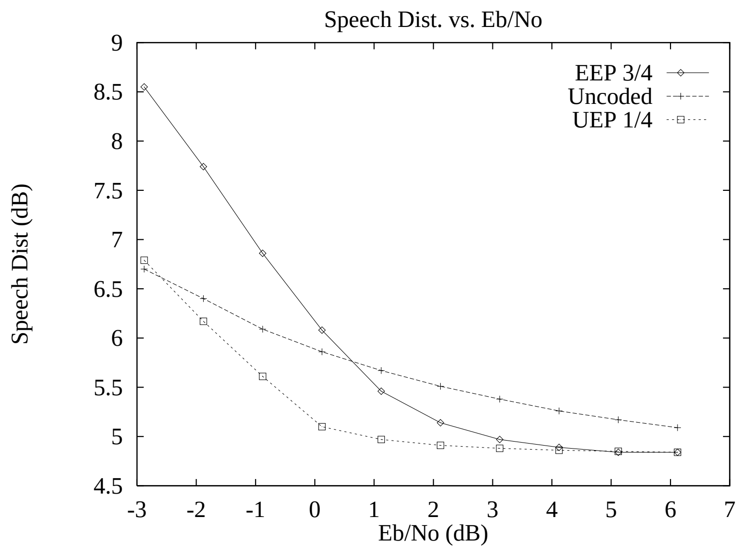

Figure 4.26: Average Speech Distortion for Rate 1/4 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

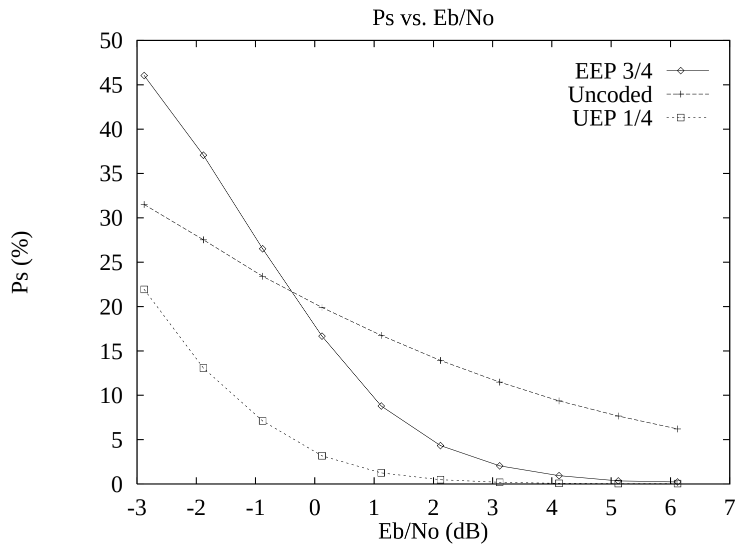

Figure 4.27: Symbol Error Rate for Rate 1/4 Unequal Error Protection and MAP2 Decoding of LSP Parameters over the Rayleigh Channel.

Chapter 5

Source Optimized Channel Decoding and Unequal Error Protection (UEP) of All the CELP parameters

As there is a variance in importance within the LSP parameters of CELP 1016, there also exists a variance in importance between the different CELP parameters. Thus, the soft decision unequal error protection schemes of the previous chapter can be applied to the entire CELP 1016 system.

5.1 Modeling the CELP Codebook Parameters

Before we define an overall system for source-optimized channel decoding and unequal protection of all the Federal Standard CELP 1016 parameters, we must first model the redundancy in the CELP codebook (CB) parameters. This includes the pitch gain and delay as well as the codebook gains and index.

Although all these parameters are of different bit-lengths, we will only model the three most significant bits of each parameter. The reasons for this simplification can be summarized as follows:

- More decoder memory is required to store the transitional probabilities of each entire parameter. For example, for the 9-bit codebook index and MAP1 decoding, a 512 by 512 matrix must be used for each of the four codebook indices.

- For unequal error protection, the modified Viterbi algorithm, which waits three regular trellis steps before computing the metric, now must wait a variable number of steps depending on the length of the parameter it is decoding. This increases the number of trellis comparisons per node, and hence the decoder complexity.

- Equal error protection as previously defined cannot be employed, since a rate k/n decoder can only use MAP decoding on parameters that are a multiple of k. For example, our rate 3/4 decoder will not be able to use MAP decoding on the 5-bit gain parameters.

Hence, in order to compare unequal error protection schemes to equal error protection schemes of similar complexity, we will use 3-bit models for all the codebook parameters. Note that codebook parameters are different from the LSP parameters in that there are 4 per frame, or 1 per sub-frame.

We let the random process, $\{U_{i,j} : 1 \leq i \leq 4, j = 1, 2, ...\}$, represent the three most significant bits of the i^{th} quantized CB parameter in frame j. Thus, $\mathbf{U}_j = [U_{1,j}, ..., U_{4,j}]$ is a vector consisting of the 4 quantized CB parameters (3 MSB's) in frame j.

We use the same two Markov models of Section 3.1.3 to quantify both the intra-frame redundancy and the inter-intra-frame redundancy of the CB parameters. The assumption is that the adaptive smoothing efforts of the CELP encoder produce a certain amount of redundancy within the CB parameters. Thus, re-expressing equation (3.38) for the CB parameters we get, the redundancy due to non-uniformity,

ρ_D, for both model A and B to be:

$$\rho_D^{A\ or\ B} = 12 + \sum_{i=1}^{4} E_{Pr(u_{i,j})}[log_2\ Pr(u_{i,j})]. \tag{5.1}$$

Note, $log_2|\mathcal{X}| = 12$ now that 12 bits are being coded every frame, instead of 30 for the LSP parameters.

Also the redundancy due to memory, ρ_M, can be rewritten for the codebook parameters (c.f. equations (3.40) and (3.42)):

$$\begin{aligned} \rho_M^{(A)} &= -\sum_{i=1}^{4} E_{Pr(u_{i,j})}[log_2\ Pr(u_{i,j})] + \\ &\quad E_{P_A^{(1)}(u_{1,j})}[log_2\ P_A^{(1)}(u_{1,j})] + \\ &\quad \sum_{i=2}^{4} E_{P_A^{(i)}(u_{i,j},u_{i-1,j})}[log_2\ P_A^{(i)}(u_{i,j}|u_{i-1,j})], \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} \rho_M^{(B)} &= -\sum_{i=1}^{4} E_{Pr(u_{i,j})}[log_2\ Pr(u_{i,j})] + \\ &\quad E_{P_B^{(1)}(u_{1,j},u_{1,j-1})}[log_2\ P_B^{(1)}(u_{1,j}|u_{1,j-1})] + \\ &\quad \sum_{i=2}^{4} E_{P_B^{(i)}(u_{i,j},u_{i-1,j},u_{i,j-1})}[log_2\ P_B^{(i)}(u_{i,j}|u_{i-1,j}, u_{i,j-1})], \end{aligned} \tag{5.3}$$

for Model A and B respectively. $P_A^{(i)}$ and $P_B^{(i)}$ are given by equations (3.22) and (3.24), respectively, except that now $1 \leq i \leq 4$.

Note that Model A and B are slightly different for the pitch delay parameters, since the odd sub-frame pitch delays and the even sub-frame pitch delays are different in nature. Hence, for both Models we do not condition on the immediately preceding parameter. Thus, Model A for the pitch delays consists only of the redundancy due to non-uniformity, ρ_D, as given by equation (5.2), and Model B consists of the redundancy due to non-uniformity, ρ_D, and the inter-frame redundancy. Model B for the pitch delays can be summarized as follows

$$\rho_M^{(B)} = -\sum_{i=1}^{4} E_{Pr(u_{i,j})}[log_2\ Pr(u_{i,j})] +$$

$$E_{P_B^{(1)}(u_{1,j},u_{1,j-1})}[log_2\ P_C^{(1)}(u_{1,j}|u_{1,j-1})] +$$
$$\sum_{i=2}^{4} E_{P_C^{(i)}(u_{i,j},u_{i,j-1})}[log_2\ P_B^{(i)}(u_{i,j}|u_{i,j-1})], \tag{5.4}$$

where

$$P_C^{(i)}(u_{i,j}|u_{i,j-1}) \triangleq Pr(U_{i,j} = u_j|U_{i,j-1} = u_{i,j-1}), \tag{5.5}$$

for $i = 1, ..., 4$.

The training sequence (83,826 frames) from the TIMIT speech database [27] was applied to the Federal Standard CELP 1016 vocoder. For every 30ms of speech, CELP analysis was performed to arrive at 16 quantized CB parameters. The relative frequency of transitions between the values of 3 MSB's of each codebook parameter were compiled to compute the Markov transition probabilities of both models. These probabilities were used with equations (5.1) to (5.4) to compute the redundancies due to non-uniformity, ρ_D and memory, ρ_M, respectively. Tables 5.1 to 5.4 display the results. Note that results for the individual subframes as well as for the entire frame are shown.

Codebook Gain	Model A			Model B		
Redundancy	ρ_D	ρ_M	ρ_T	ρ_D	ρ_M	ρ_T
Sub Frame 1	1.0660	0.0000	1.0660	1.0660	0.1495	1.2155
Sub Frame 2	0.9612	0.3137	1.2749	0.9612	0.3601	1.3213
Sub Frame 3	1.0340	0.3631	1.3971	1.0340	0.3640	1.3980
Sub Frame 4	0.9866	0.3472	1.3337	0.9866	0.3808	1.3673
Frame Redundancy	4.0478	1.024	5.07178	4.0478	1.2544	5.3022

Table 5.1: Codebook Gain Redundancies (in Bits/Frame) using 83,826 Frames of the TIMIT Speech Database.

Pitch Gain Redundancy	Model A			Model B		
	ρ_D	ρ_M	ρ_T	ρ_D	ρ_M	ρ_T
Sub Frame 1	0.0505	0.0000	0.0505	0.0505	0.1674	0.2179
Sub Frame 2	0.0367	0.3852	0.4220	0.0367	0.4492	0.4859
Sub Frame 3	0.0453	0.3352	0.3805	0.0453	0.3987	0.4441
Sub Frame 4	0.0507	0.4131	0.4637	0.0507	0.4757	0.5264
Frame Redundancy	0.1832	1.1335	1.3167	0.1832	1.4910	1.6742

Table 5.2: Pitch Gain Redundancies (in Bits/Frame) using 83,826 Frames of the TIMIT Speech Database.

Interpretation of results:

- The codebook gain parameters have the most redundancy (44.2 % with Model B), followed by the pitch gains (14.0 % with Model B), the pitch delays (with 12.78 % with Model B) and finally the codebook indices (0.5 % with Model B).
- The majority of the redundancy in the codebook gains is due to non-unformity; i.e., for Model B $\rho_D = 4.0478$ compared to $\rho_M = 1.2544$.
- The pitch gains have almost no redundancy due to non-uniformity, $\rho_D = 0.1832$.
- The codebook indices have minimal redundancy. This is because they are chosen from a randomly generated stochastic codebook.

Intuitively, these results suggest that the codebook gains will benefit the most from MAP decoding, followed by the pitch gains and then the pitch delays. The codebook index parameter exhibit essentially no redundancy, and thus, should not be decoded with the MAP algorithm. Therefore, for our overall MAP decoding scheme,

Pitch Delay Redundancy	Model A			Model B		
	ρ_D	ρ_M	ρ_T	ρ_D	ρ_M	ρ_T
Sub Frame 1	0.0267	0.0000	0.0267	0.0267	0.2403	0.2670
Sub Frame 2	0.3165	0.0000	0.3165	0.3165	0.1152	0.4317
Sub Frame 3	0.0235	0.0000	0.0235	0.0235	0.3756	0.3991
Sub Frame 4	0.3397	0.0000	0.3397	0.3397	0.0955	0.4352
Frame Redundancy	0.7064	0.0000	0.7064	0.7064	0.8266	1.5330

Table 5.3: Pitch Delay Redundancies (in Bits/Frame) using 83,826 Frames of the TIMIT Speech Database.

we will only exploit the redundancies present in the LSP's, codebook gains, pitch gains and pitch delays.

5.2 Source Optimized Channel Decoding for all the CELP parameters

We use one general model for the overall CELP transmission system, and simulate the results using EEP rate 3/4 and 1/2, uncoded and UEP base rate 1/3 systems. In this model, only 78 bits per CELP frame* are coded and source optimized channel decoded. These bits are:

- 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,

*Recall that there are 144 bits per CELP frame.

Codebook Index Redundancy	Model A			Model B		
	ρ_D	ρ_M	ρ_T	ρ_D	ρ_M	ρ_T
Sub Frame 1	0.0053	0.0000	0.0053	0.0053	0.0018	0.0070
Sub Frame 2	0.0090	0.0067	0.0157	0.0090	0.0111	0.0201
Sub Frame 3	0.0065	0.0060	0.0125	0.0065	0.0099	0.0165
Sub Frame 4	0.0115	0.0054	0.0170	0.0115	0.0093	0.0208
Frame Redundancy	0.0323	0.0181	0.0504	0.0323	0.0321	0.0645

Table 5.4: Codebook Index Redundancies (in Bits/Frame) using 83,826 Frames of the TIMIT Speech Database.

- 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 ML decoding, for all three soft decision algorithms. The remaining 60 bits are sent uncoded and hard decision decoded for all transmission schemes. This includes all the codebook indices, since it has been shown that they contain little redundancy and are the least important to speech production. 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 testing sequence used is once again the 4753-frame (2.2 minute) TIMIT sequence, half uttered by females and half uttered by males. The performance criteria used are:

- The average speech distortion measure, (see Appendix A) which has a minimum possible distortion of 4.79 dB.

- The symbol error rate, P_s, which now includes both the coded and uncoded bit errors of all the parameters.

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

The average spectral distortion measure is not used since it only measures the spectral characteristics of the speech, and thus, is only relevant to the LSP parameters.

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.

We next propose the various transmission schemes, and evaluate their performance.

5.2.1 Various Equal Error Protection Transmission Schemes

The first transmission scheme is based on the 32-state, rate 3/4 convolutional coder of Section 3.2.1 (see equation (3.43)). The transmission system outlined in section 3.2 is employed. It uses the three soft decision decoding algorithms, ML, MAP1, and MAP2, outlined by equations (3.51), (3.52), and (3.53), respectively. The following modified version of equation (3.53) is used to take the larger sequence frame size into consideration

$$\sum_{k=1}^{K} \parallel \mathbf{y}_k - \mathbf{a}_k \hat{\mathbf{x}}_k \parallel^2 \; - \; N_0 \; ln \; P_B^{([k \bmod 10])}(\mathbf{u}_k | \mathbf{u}_{k-1}, \mathbf{u}_{k-26}), \tag{5.6}$$

where, all the equation parameters are the same as in Chapter 3. Note, the change to u_{k-26} from u_{k-10} since 26 three-bit symbols are transmitted per frame. Once again decoder delay of 1 frame is employed, which now corresponds to 26 symbols, thus, the path memory $p = 26 = 13m$.

Simulations were performed for both the AWGN and Rayleigh fading channels, and the results are summarized in Tables 5.5 and 5.6.

The second transmission system is based on the rate 1/2 mother code of the RCPC family of codes described in Section 4.3.1, where

$$\mathbf{G}(D) = [1 + D^2 + D^4 + D^5 \quad 1 + D^1 + D^2 + D^3 + D^5]. \tag{5.7}$$

The system employed is the one outlined in Section 4.2.1. The soft decision decoding algorithms for ML and MAP1 decoding are given by equations (4.10) and (4.11), respectively. For MAP2 decoding, equation (4.12) can be rewritten as follows

$$\sum_{k=1}^{K}\sum_{i=1}^{3}(\| \mathbf{y}_{k,i} - \mathbf{a}_{k,i}\hat{\mathbf{x}}_{k,i} \|^2) \; - \; N_0 \; ln \; P_B^{([k \bmod 10])}(\mathbf{u}_k|\mathbf{u}_{k-1}, \mathbf{u}_{k-26}), \tag{5.8}$$

where, all the parameters are the same as before. The path memory corresponding to a one subframe delay in this system is $p = 78 = 15.6m$. The simulations are shown in Tables 5.7 and 5.8.

The final equal error protection system is one without any channel coding. It is completely outlined in Section 4.2.3, with the following modification to the MAP2 metric

$$\sum_{k=1}^{K}\sum_{i=1}^{3}(\| y_{k,i} - a_{k,i}\hat{x}_{k,i} \|^2) \; - \; N_0 \; ln \; P_B^{([k \bmod 10])}(\mathbf{u}_k|\mathbf{u}_{k-1}, \mathbf{u}_{k-26}). \tag{5.9}$$

As previously mentioned, these simulations are performed frame by frame; the decoding delay is only one frame. The simulations results for both channel criteria are summarized in Tables 5.9 and 5.10.

–	Speech Dist. (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-4.00 dB	12.89	10.94	10.63	89.74 %	68.06 %	67.60 %
-3.00 dB	12.68	9.51	9.47	87.70 %	60.82 %	59.60 %
-2.00 dB	12.26	8.29	7.98	84.34 %	51.26 %	48.74 %
-1.00 dB	11.53	7.09	6.73	76.72 %	39.28 %	36.80 %
0.00 dB	9.65	6.18	5.91	59.78 %	26.50 %	24.91 %
1.00 dB	6.99	5.56	5.49	33.63 %	16.92 %	16.53 %
2.00 dB	5.60	5.33	5.33	14.19 %	11.01 %	11.04 %

Table 5.5: Speech Distortion and Percentage Symbol Error for Rate-3/4 Equal Error Protection of CELP Parameters over the AWGN Channel.

–	Speech Dist. (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.00 dB	13.32	11.24	10.90	90.65 %	69.95 %	69.71 %
-2.00 dB	13.08	10.14	10.00	89.43 %	64.67 %	63.80 %
-1.00 dB	12.87	9.05	8.80	87.97 %	58.25 %	56.25 %
0.00 dB	12.47	8.05	7.72	85.13 %	50.31 %	47.81 %
1.00 dB	11.92	7.12	6.73	80.57 %	41.53 %	38.97 %
2.00 dB	10.97	6.40	6.17	72.36 %	32.82 %	30.93 %
3.00 dB	9.26	5.87	5.78	58.74 %	24.96 %	24.16 %
4.00 dB	7.44	5.60	5.52	40.50 %	19.32 %	18.87 %
5.00 dB	6.16	5.41	5.39	24.62 %	15.41 %	15.30 %
6.00 dB	5.57	5.33	5.32	15.66 %	12.64 %	12.64 %

Table 5.6: Speech Distortion and Percentage Symbol Error for Rate-3/4 Equal Error Protection of CELP Parameters over the Rayleigh Fading channel.

–	Speech Dist. (dB)			P_s(%)		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-4.00 dB	12.32	9.25	8.90	82.89 %	61.40 %	59.58 %
-3.00 dB	11.30	7.98	7.60	74.68 %	52.12 %	49.85 %
-2.00 dB	9.53	6.87	6.47	60.75 %	41.16 %	39.33 %
-1.00 dB	7.60	6.04	5.86	42.69 %	31.33 %	30.24 %
0.00 dB	6.14	5.65	5.61	27.84 %	23.89 %	23.62 %
1.00 dB	5.59	5.47	5.47	19.63 %	18.93 %	18.93 %
2.00 dB	5.37	5.35	5.36	14.92 %	14.85 %	14.88 %

Table 5.7: Speech Distortion and Percentage Symbol Error for Rate-1/2 Equal Error Protection of CELP Parameters over the AWGN Channel.

–	Speech Dist. (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.00 dB	12.68	9.27	9.00	84.35 %	62.13 %	60.30 %
-2.00 dB	11.99	8.14	7.81	77.84 %	54.22 %	52.02 %
-1.00 dB	10.48	7.20	6.81	67.72 %	45.29 %	43.09 %
0.00 dB	8.61	6.41	6.15	53.61 %	36.87 %	35.57 %
1.00 dB	7.10	5.89	5.76	39.15 %	30.31 %	29.47 %
2.00 dB	6.11	5.65	5.64	29.04 %	25.62 %	25.42 %
3.00 dB	5.65	5.53	5.54	23.14 %	22.24 %	22.22 %
4.00 dB	5.47	5.45	5.46	19.62 %	19.46 %	19.49 %
5.00 dB	5.41	5.40	5.41	17.04 %	17.00 %	17.05 %
6.00 dB	5.35	5.35	5.36	14.70 %	14.70 %	14.75 %

Table 5.8: Speech Distortion and Percentage Symbol Error for Rate-1/2 Equal Error Protection of CELP Parameters over the Rayleigh Fading Channel.

–	Speech Dist. (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-4.00 dB	9.79	8.23	8.16	63.87 %	55.40 %	54.55 %
-3.00 dB	9.35	7.65	7.49	57.77 %	49.37 %	48.43 %
-2.00 dB	8.71	7.18	6.97	50.69 %	42.82 %	41.92 %
-1.00 dB	8.22	6.73	6.59	42.87 %	35.73 %	34.77 %
0.00 dB	7.58	6.41	6.21	34.54 %	28.58 %	27.62 %
1.00 dB	7.01	6.08	5.94	26.08 %	21.36 %	20.69 %
2.00 dB	6.51	5.76	5.67	18.21 %	14.80 %	14.35 %

Table 5.9: Speech Distortion and Percentage Symbol Error for Uncoded CELP Parameters over the AWGN Channel.

–	Speech Dist. (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.00 dB	10.27	8.45	8.33	68.50 %	58.82 %	57.89 %
-2.00 dB	9.91	8.01	7.78	64.33 %	54.52 %	53.50 %
-1.00 dB	9.49	7.59	7.33	59.71 %	50.10 %	48.87 %
0.00 dB	9.12	7.23	7.00	54.66 %	45.45 %	44.25 %
1.00 dB	8.75	6.93	6.71	49.38 %	40.69 %	39.41 %
2.00 dB	8.30	6.67	6.48	44.04 %	36.03 %	34.91 %
3.00 dB	8.00	6.43	6.25	38.78 %	31.45 %	30.46 %
4.00 dB	7.70	6.23	6.07	33.77 %	27.18 %	26.27 %
5.00 dB	7.29	6.05	5.91	29.13 %	23.31 %	22.48 %
6.00 dB	6.96	5.92	5.80	24.67 %	19.73 %	19.09 %

Table 5.10: Speech Distortion and Percentage Symbol Error for Uncoded CELP Parameters over the Rayleigh Fading Channel.

5.2.2 Base Rate 1/3 Unequal Error Protection Transmission Scheme

The base rate-1/3 family of RCPC codes [14] is selected for the unequal error protection system for transmitting all the CELP parameters. This is because it provides better results than the rate-1/2 family of codes [20]. The rate-1/4 RCPC code family was not used since it was a 16-state system, and is not consistent with our two EEP systems of rate 3/4 and 1/2, respectively.

The system employed is the one outlined in Section 4.2.1. The soft decision decoding algorithms for ML, MAP1 and MAP2 decoding are given by equations (4.10), (4.11), and (5.8), respectively. The protection scheme for the CELP parameters is consistent with the findings of Section 2.5, where the LSP's are given the most protection, followed by the odd subframe pitch delays, the even subframe pitch delays, and finally the codebook gains. The pitch gains are sent uncoded with respective soft decision decoding in the case of MAP1 and MAP2 transmission.

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 5.11: Overall CELP Unequal Error Protection Scheme using Rate-1/3 Family of RCPC codes.

The simulations are performed using a decoding delay of one frame in length

($p = 78 = 15.6m$). The simulations results for both channel criteria are summarized in Tables 5.12 and 5.13.

5.2.3 Performance Comparison of Unequal Error Protection System versus Equal Error Protection Systems

Tables 5.14 to 5.17 show the coding gains for the same value of average speech distortion for the UEP system versus both the EEP rate 3/4 and EEP rate 1/2. Figures 5.1 and 5.2 depict the different transmission schemes for MAP2 decoding over both the AWGN and Rayleigh fading channels.

Observations Regarding Results:

- At an average speech distortion of 6.0 dB, the gains for UEP versus EEP rate-3/4 (both MAP2) are 1.42 dB and 2.55 dB, for 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.

- At a speech distortion of 6.0 dB, the coding gains for UEP versus the lower rate 1/2 EEP are 0.3 dB and 0.5 dB, for the AWGN and Rayleigh channels. Figures 5.1 and 5.2 also graphically show that the UEP system is performing better than the EEP rate 1/2 system.

- The amount of gain due to UEP compared to gain solely due to MAP decoding is relatively higher for the overall system. For instance, at an average speech distortion of 6.5 dB, for UEP vs. EEP rate 3/4 (see Table 5.14), the gain solely from UEP is 2.21 dB compared to the gain for MAP2 which is 2.07 dB. Similarly, for the same channel conditions and RCPC family of codes but coding only the LSP's, the gain only due to UEP is 1.10 dB compared to 3.51 dB for MAP decoding (see Table 4.25). This is partly due to the fact that other CELP

–	Speech Dist. (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-4.00 dB	10.53	7.62	7.30	73.30 %	55.70 %	53.75 %
-3.00 dB	8.82	6.77	6.53	61.31 %	47.39 %	45.80 %
-2.00 dB	7.30	6.19	6.09	47.59 %	39.53 %	38.67 %
-1.00 dB	6.34	5.90	5.90	36.61 %	33.37 %	32.99 %
0.00 dB	6.01	5.79	5.77	29.72 %	28.66 %	28.53 %
1.00 dB	5.90	5.71	5.71	25.01 %	24.47 %	24.41 %
2.00 dB	5.79	5.63	5.62	20.61 %	20.16 %	20.15 %

Table 5.12: Speech Distortion and Percentage Symbol Error for Base Rate-1/3 Unequal Error Protection of CELP Parameters over the AWGN Channel.

–	Speech Dist. (dB)			$P_s(\%)$		
$\frac{E_b}{N_0}$	ML	MAP1	MAP2	ML	MAP1	MAP2
-3.00 dB	10.38	7.42	7.12	73.43 %	55.18 %	53.09 %
-2.00 dB	8.82	6.71	6.47	62.58 %	48.21 %	46.51 %
-1.00 dB	7.38	6.24	6.16	50.65 %	41.71 %	40.74 %
0.00 dB	6.55	6.03	5.98	40.65 %	36.90 %	36.27 %
1.00 dB	6.17	5.88	5.87	34.69 %	32.99 %	32.86 %
2.00 dB	6.01	5.81	5.80	30.84 %	29.92 %	29.83 %
3.00 dB	5.98	5.77	5.76	27.70 %	26.99 %	26.89 %
4.00 dB	5.90	5.68	5.67	24.85 %	24.15 %	24.09 %
5.00 dB	5.81	5.61	5.62	21.94 %	21.29 %	21.27 %
6.00 dB	5.71	5.55	5.56	19.23 %	18.62 %	18.59 %

Table 5.13: Speech Distortion and Percentage Symbol Error for Base Rate-1/3 Unequal Error Protection of CELP Parameters over the Rayleigh Fading Channel.

parameters do not have as much redundancy as is exhibited in the LSP's, and partly due to the improved performance of unequal error protection for all the CELP parameters.

- Note that in Table 5.8 that the speech distortion for the ML and MAP2 case are essentially equal at a E_b/N_0 over 4 dB. This is because this is a more powerful code, and thus, MAP becomes advantageous at lower SNR's.

5.2.4 Subjective Listening Tests

The final test is a subjective one, since speech recognition is quite subjective in nature. We performed listening tests that made pairwise comparisons between three different transmission systems at two different E_b/N_0's. The systems we implemented used MAP2 decoding; they consist of: rate-3/4 EEP, rate-1/2 EEP and UEP using the rate-1/3 RCPC family. 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*. This way, six pairwise comparisons were made at each E_b/N_0 by each listener, resulting in 100 trials for each test and at each signal to noise ratio (E_b/N_0).

The results of the tests are shown in Table 5.18.

Observations Regarding Results:

- At low SNR, $E_b/N_0 = -2$ dB, the UEP scheme performs significantly better than both EEP schemes for the first speech segment, with 70% choosing UEP

Coding Gains	Avg. Speech Dist. (dB)			
	5.5	6.0	6.5	7.0
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+1.10	+1.82	+2.07	+2.21
MAP2 – UEP 1/3 vs. MAP2 – EEP 3/4	-2.36	+1.42	+2.21	+2.39
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 3/4	-1.26	+3.24	+4.28	+4.61

Table 5.14: 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)			
	5.5	6.0	6.5	7.0
MAP2 – EEP 3/4 vs. ML – EEP 3/4	+1.48	+2.47	+3.20	+3.59
MAP2 – UEP 1/3 vs. MAP2 – EEP 3/4	-2.56	+2.55	+3.46	+3.54
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 3/4	-1.08	+5.02	+6.66	+7.14

Table 5.15: Coding Gains for Base Rate 1/3 Unequal Error Protection over Rate 3/4 Equal Error Protection for the Rayleigh Fading channel.

Coding Gains	Avg. Speech Dist. (dB)			
	5.5	6.0	6.5	7.0
MAP2 – EEP 1/2 vs. ML – EEP 1/2	+0.62	+1.48	+1.78	+1.88
MAP2 – UEP 1/3 vs. MAP2 – EEP 1/2	-2.55	+0.30	+0.91	+1.14
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 1/2	-1.92	+1.78	+2.69	+3.02

Table 5.16: Coding Gains for Base Rate 1/3 Unequal Error Protection over Rate 1/2 Equal Error Protection for the AWGN channel.

Coding Gains	Avg. Speech Dist. (dB)			
	5.5	6.0	6.5	7.0
MAP2 – EEP 1/2 vs. ML – EEP 1/2	0.00	+1.85	+2.14	+2.29
MAP2 – UEP 1/3 vs. MAP2 – EEP 1/2	-3.21	+0.50	+1.52	+1.63
Total Gain MAP2 – UEP 1/3 vs. ML – EEP 1/2	-3.29	+2.35	+3.65	+3.92

Table 5.17: Coding Gains for Base Rate 1/3 Unequal Error Protection over Rate 1/2 Equal Error Protection for the Rayleigh Fading channel.

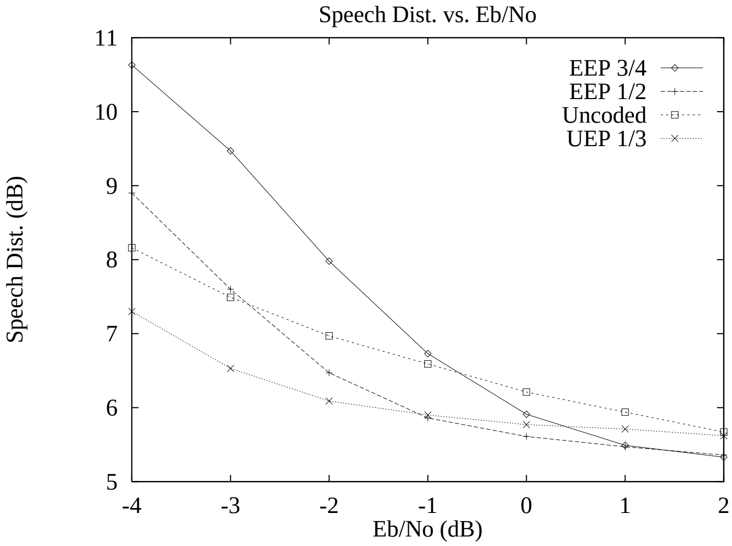

Figure 5.1: Average Speech Distortion for Rate 1/3 Unequal Error Protection and MAP2 Decoding of all the CELP parameters over the AWGN Channel.

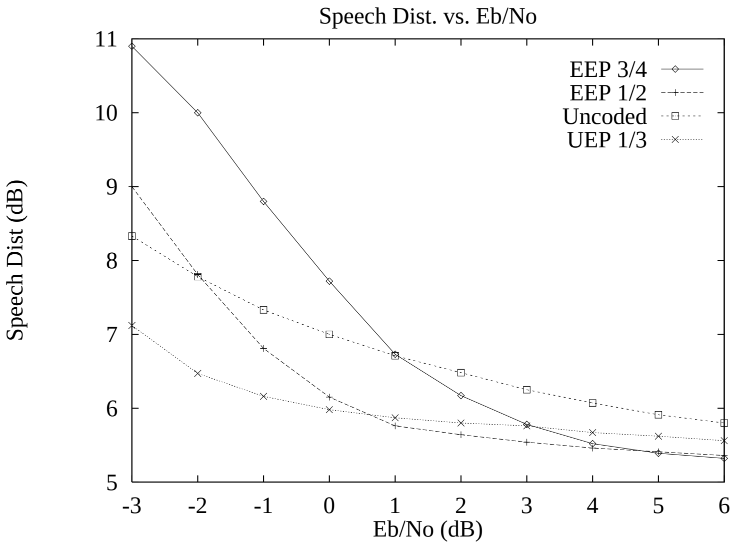

Figure 5.2: Average Speech Distortion for Rate 1/3 Unequal Error Protection and MAP2 Decoding of all the CELP parameters over the Rayleigh Channel.

E_b/N_0 = -2 dB	Speech Segment 1	EEP 1/2 vs. EEP 3/4	EEP 1/2: 92 % EEP 3/4: 2% Neither: 6%
		UEP vs. EEP 3/4	UEP: 96 % EEP 3/4: 0% Neither: 4%
		UEP vs. EEP 1/2	UEP: 70% EEP 1/2: 2% Neither: 28%
E_b/N_0 = -2 dB	Speech Segment 2	EEP 1/2 vs. EEP 3/4	EEP 1/2: 74 % EEP 3/4: 2% Neither: 24%
		UEP vs. EEP 3/4	UEP: 84 % EEP 3/4: 0% Neither: 16%
		UEP vs. EEP 1/2	UEP: 40% EEP 1/2: 8% Neither: 52%
E_b/N_0 = 1 dB	Speech Segment 3	EEP 1/2 vs. EEP 3/4	EEP 1/2: 98% EEP 3/4: 2% Neither: 0%
		UEP vs. EEP 3/4	UEP: 70% EEP 3/4: 14% Neither: 16%
		UEP vs. EEP 1/2	UEP: 16% EEP 1/2: 60% Neither: 24%
E_b/N_0 = 1 dB	Speech Segment 4	EEP 1/2 vs. EEP 3/4	EEP 1/2: 74% EEP 3/4: 8% Neither: 18%
		UEP vs. EEP 3/4	UEP: 90% EEP 3/4: 0% Neither: 10%
		UEP vs. EEP 1/2	UEP: 36% EEP 1/2: 18% Neither: 46%

Table 5.18: Listening Test Results for Unequal Error Protection versus Equal Error Protection over the Rayleigh Fading Channel.

over EEP rate 1/2, and over 90% selecting UEP over EEP rate 3/4.

- The second speech segment, at $E_b/N_0 = -2$ dB, did not provide as clear a result as the first for the case of UEP versus the EEP rate 1/2 scheme. Even though 40 % selected UEP over EEP rate 1/2, 52 % chose neither. The UEP scheme still outperformed the EEP rate 3/4 for speech segment 2.

- At the higher SNR of $E_b/N_0 = 2$ dB, the UEP scheme still outperforms the EEP rate 3/4. However, the EEP rate 1/2 system begins to perform better than the UEP scheme, and actually beat the UEP system for speech segment 3.

5.2.5 Conclusions

Objectively, it is evident that the unequal error protection scheme outperforms the equal error protection schemes at low E_b/N_0. For the AWGN channel the UEP scheme begins to outperform the EEP rate-1/2 at around $E_b/N_0 = -1$ dB, while this crossover occurs at around 1 dB for the Rayleigh fading channel.

Subjectively, the UEP coding system beat the two EEP systems a total of three out of four times in listening tests at low SNR's, with the fourth test being inconclusive. However, at higher SNR's the UEP scheme could only beat the EEP systems two out of four times, and was actually outperformed by the EEP rate 1/2 system in one trial.

The numerical data shows that at high SNR's the equal error coding schemes approach an asymptotically lower error rate than the unequal protection schemes. Thus, a combination of the two schemes, with EEP at high SNR's and UEP at low SNR's would produce an overall robust system that degrades smoothly during severe channel conditions.

Chapter 6

Conclusion

6.1 Summary

In this work, we investigated the design and implementation of joint source channel coding and unequal error protection schemes to reliably transmit CELP 1016 encoded speech over very noisy channels. This included both the AWGN channel and the fully interleaved Rayleigh fading channel both used in conjunction with BPSK modulation.

In Chapter 2, we discussed the theory behind CELP speech coding, and summarized the various differences in sensitivity of the CELP parameters to channel noise.

In Chapter 3, we characterized both inter-frame and intra-frame redundancies in the CELP LSP parameters, and employed the Viterbi MAP metric for soft decision decoding at the output. We simulated transmission over both the AWGN and fully interleaved Rayleigh fading channels. We achieved coding gains of up to 5 dB in E_b/N_0 under the spectral distortion criterion via MAP-EEP decoding schemes that exploited the residual redundancy present in the LSP parameters.

In Chapter 4, we used unequal error protection in conjunction with MAP decoding for the LSP parameters. We found that the unequal error protection schemes increased the error resiliency of the system at low values of E_b/N_0. Coding gains of up to 2.5 dB for the same average speech distortion were realized with rate-1/4 unequal error protection over rate-3/4 equal error protection.

Chapter 5 examined the redundancy present in all the CELP parameters. It was observed that the three most significant bits of the codebook gains actually exhibited more redundancy than the three-bit LSP symbols, at around 44%. The other parameters exhibited significantly less redundancy, with the 3 MSB's of the codebook indices showing no significant redundancy.

We then exploited the redundancy in all the CELP 1016 parameters through the MAP Viterbi metric and attained coding gains in E_b/N_0 for the same speech distortion of 2.2 dB and 3.6 dB, for the AWGN and Rayleigh fading channels, respectively.

An overall unequal error protection scheme was then implemented through a rate-1/3 RCPC family of codes. We achieved coding gains of up to 3.5 dB and 1.6 dB over EEP rate 3/4 and rate 1/2 schemes. Subjective listening tests were also performed, and showed that at $E_b/N_0 = -2$ dB for the Rayleigh fading channel the UEP scheme did perform better than the EEP schemes.

The thesis shows that unequal error protection of CELP 1016 encoded speech over very noisy channels performs sginificantly better than equal error protection schemes of the same source encoded speech.

6.2 Future Work

It is clear that unequal error protection of the Federal Standard CELP 1016 parameters does provide significant gains in performance at low SNR's. However, at high SNR's equal error protection channel coding schemes perform better. Thus, an *adaptive* system that estimates the channel error conditions and uses this information to select an appropriate error protection scheme – UEP at low SNR's and EEP at high

SNR's – would result in a *robust* transmission system. Recall that CELP 1016 does have a expansion bit that could be used for this purpose.

One assumption in this thesis is that of an ideal infinite interleaver. This is not realizable in a real world system. Thus, the unequal error protection systems of this paper could be implemented without this assumption. This type of work was done by Hagenaur *et al.* in [16], and it was shown that a finite interleaver did not significantly affect the performance of the unequal error protection system.

This study could also be extended to other similar vocoder designs. Although CELP 1016 is a standard, many newer more efficient linear predictive voice coders have been designed. These exhibit less redundancy in the source; it would be interesting to see how the protection schemes explored in this thesis would apply to these new vocoders.

Another interesting extension would be to optimize the channel codes (in this case the RCPC codes) for the source characteristics. This source optimized channel coding method was explored in [18].

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 [3]. In this method different signal energy levels are allocated relative to the importance of that specific signal in the frame. This method can be applied to our system in conjunction with the RCPC methods of unequal error protection.

Appendix A

Derivation of Average Speech Distortion Measure

The average speech distortion measure first discussed in Section 3.2.3 consists of seven different weighted distance measures of two different types – cosh measures and cepstral distance [12]. The parameters needed to calculate these measures will be discussed followed by the definitions of these measures, and finally the definition of the average speech distortion measure.

A.1 Parameters Needed For Calculating Distance Measures

A.1.1 Autocorrelation Representation of Speech Sequences

Evaluation of all the distance measurements can be done efficiently through the use of autocorrelation sequences [17]. The autocorrelation of a discrete-time deterministic signal is defined as [29]

$$\phi(i,k) \triangleq \sum_{m=-\infty}^{\infty} x(m-i)x(m-k). \tag{A.1}$$

Since we will be dealing with short-term spectra, we can further define a short-term autocorrelation function as [29]

$$r(k) \triangleq \sum_{m=-\infty}^{\infty} [x(m)w(m)][x(m+k)w(m+k)]^*, \tag{A.2}$$

where $w(m)$ is a window function such that $w(m) = w(-m)$. Equation (A.2) states that the input samples, $\{x(m)\}$, are multiplied by a window, $w(m)$, that selects a short fragment of speech equal in length to the length of the window.

For our distance measures, a Hamming window of length N is used to weight all the samples in the speech sequence. The Hamming window is defined by

$$w_h(m) = 0.54 - 0.46cos\left(\frac{2\pi m}{(N-1)}\right), \quad 0 \leq n \leq N-1. \tag{A.3}$$

The length $N = 60$ (or 7.5 ms) in our case since we will be evaluating the distance criteria over CELP subframes.

Since our window, $w_h(m)$, is of a finite duration, equation (A.2) can be re-expressed as

$$r(k) = \sum_{m=0}^{N-1-k} [x(m)w_h(m)][x(m+k)w_h(m+k)], \quad k = 0, 1, 2, ..., \tag{A.4}$$

where, the term $r(k)$ is referred to as the k^{th} autocorrelation lag. In CELP 1016 we use 10^{th} order autocorrelation analysis, which means we will calculate up to the 10^{th} autocorrelation lag (ie. $k = 1, 2, ..., 10$).

We represent autocorrelation lags of the input speech sequence, $\{x(m)\}$, and the output speech sequence, $\{x^{'}(m)\}$, autocorrelation lags by $r_x(k)$ and $r^{'}_x(k)$, respectively.

$*$In all the following derivations, we assume that the speech samples in question start at time $m = 0$. This is true only for the first segment of speech, and thus all other segments must be shifted accordingly for the following calculations to be applicable.

A.1.2 Estimating the LP Inverse Filter Coefficients

Recall that the inverse all-pole filter used in LP analysis is of the form

$$H(z) = \frac{\sigma}{A(z)}, \tag{A.5}$$

where,

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}, \tag{A.6}$$

and, $\{a_k\}_{k=1,\ldots,p}$ are the *actual* filter coefficients, $a_0 = 1$, and p is the filter order ($p = 10$ in the CELP 1016 vocoder). The gain of the filter is represented by σ.

For the all-pole linear predictive system of Figure 2.1, the speech sample $x(m)$ is related to the excitation $u(m)$ by the following equation

$$x(m) = \sum_{k=1}^{p} a_k x(m-k) + \sigma u(m). \tag{A.7}$$

If the *predicted* LP coefficients, denoted $\{\alpha_k\}_{k=1,\ldots,p}$, produce a sequence of speech samples, $\{\hat{x}(m)\}$, then the prediction error, $e(m)$, can be defined

$$e(m) \triangleq x(m) - \hat{x}(m) = x(m) - \sum_{k=1}^{p} \alpha_k x(m-k). \tag{A.8}$$

The short-term mean squared prediction error, E, can now be defined as

$$E \triangleq \sum_{m=-\infty}^{\infty} e^2(m) \tag{A.9}$$

$$= \sum_{m=-\infty}^{\infty} \left(x(m) - \sum_{k=1}^{p} \alpha_k x(m-k) \right). \tag{A.10}$$

By setting $\partial E / \partial \alpha_i = 0$ for $i = 1, 2, ..., p$, we can find the values of α_k that minimize the average prediction error. Using equation (A.1), we can obtain the following set of equations

$$\sum_{k=1}^{p} \alpha_k \phi(i,k) = \phi(i,0), \quad i = 1, 2, ..., p. \tag{A.11}$$

This set of p linear equations can be solved, using the autocorrelation method [8] [29], to produce the *unknown* predictor LP coefficients $\{\alpha_k\}_{k=1,\ldots,p}$ that minimize E for the speech segment $\{x(m)\}$ over $0 \leq m \leq N-1$.

A.1.2.1 The Autocorrelation Method for Finding the LP Inverse Filter Coefficients

In this method of solving the LP equations given by (A.11), we first use a N^{th} order Hamming window, $w_h(m)$, on the speech samples. Thus, $s(m) = x(m)w_h(m)$ over the interval $0 \leq m \leq N-1$. Note, that our window length corresponds to one subframe, therefore, $N = 60$.

Hence, the mean squared prediction error, E, will be defined over the interval $0 \leq m \leq N-1+p$, where p is the LP filter order. So E can be rewritten

$$E = \sum_{m=0}^{N+p-1} e^2(m). \tag{A.12}$$

Over this finite interval $\phi(i,k)$ can be shown to simplify to

$$\phi(i,k) \triangleq \sum_{m=0}^{N-1-(i-k)} s(m)s(m+i-k), \tag{A.13}$$

for $1 \leq i \leq p$ and $0 \leq k \leq p$. This can now be expressed in terms of short-term autocorrelation as in equation (A.4) evaluated for $(i-k)$. In other words $\phi(i,k) = r(i-k)$. Also, since correlation functions are even by definition,

$$\phi(i,k) = r(|i-k|), \tag{A.14}$$

for $1 \leq i \leq p$ and $0 \leq k \leq p$. Thus, equation (A.11) can be written as

$$\sum_{k=1}^{p} \alpha_k r(|i-k|) = r(i), \quad i = 1, 2, ..., p. \tag{A.15}$$

These are called the *autocorrelation equations.* Thus, we can use the autocorrelation of the input speech sequence, $r_x(k)$, to determine the LP coefficients, $\{\alpha_k\}$, of the input inverse filter $A(z)$. Likewise, we can determine $\{\alpha_k^{'}\}$ for the output speech filter $A^{'}(z)$ through the autocorrelation sequence $r_x^{'}(k)$. One method for solving these equations is the Durbin recursive solution [8] [29].

A.1.2.2 Durbin's Recursive Solution for the Autocorrelation Equations

The most efficient method for solving (A.15) is the Durbin recursive method, which can be stated as follows

$$E^{(0)} = r(0) \tag{A.16}$$

$$\alpha_0 = 1 \tag{A.17}$$

$$k_i = \frac{-\left[r(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} r(i-j)\right]}{E^{(i-1)}}, \quad 1 \leq i \leq p \tag{A.18}$$

$$\alpha_i^{(i)} = k_i \tag{A.19}$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \tag{A.20}$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)}. \tag{A.21}$$

Note, that $\{k_i\}$ represents the PARCOR coefficients also known as the reflection coefficients (See Section 3.1.1).

The above system of equations can be solved for $i = 1, ..., 10$ in our case, where the final LP coefficients are given by

$$\alpha_j = \alpha_j^{(10)} \quad 1 \leq j \leq 10. \tag{A.22}$$

.

A.1.3 Autocorrelation Representation of LP Inverse Filter Coefficients

We also need in our computations the autocorrelation sequences of the the LP inverse filter coefficient. Let $r_a(k)$ be the autocorrelation sequence for the input inverse filter $A(z)$, and likewise let $r_a^{'}(k)$ be the autocorrelation sequence for the output filter $A^{'}(z)$.

We now derive the filter autocorrelation sequences, $r_a(k)$ and $r_a^{'}(k)$ [12] from their respective LP filter coefficients, $\{\alpha_k\}_{k=1,...,10}$ and $\{\hat{\alpha}_k\}_{k=1,...,10}$. This is done through the definition of short-term autocorrelation, where [29]

$$r_a(k) = \sum_{m=0}^{N-i} s(m)s(m+i), \quad 1 \leq i \leq p, \tag{A.23}$$

and

$$r_a^{'}(k) = \sum_{m=0}^{N-i} s^{'}(m)s^{'}(m+i), \quad 1 \leq i \leq p. \tag{A.24}$$

A.1.4 Cepstral Representation of LP Inverse Filter Coefficients

Cepstral coefficients are another way of representing the LP filter coefficients. In particular, for a p^{th} order all-pole minimum phase filter,

$$\frac{1}{A_p(z)}, \tag{A.25}$$

a Taylor series expansion shows that [12]

$$ln[A_p(z)] = -\sum_{k=1}^{\infty} c_k z^{-k}, \tag{A.26}$$

where $\{c_k\}_{k=1,...,p}$ are the *cepstral coefficients*. Taking the derivative of both sides with respect to z^{-1} produces

$$\sum_{k=1}^{p} k\alpha_k z^{-k} = -\left[\sum_{k=1}^{p} kc_k z^{-k}\right]\left[\sum_{k=0}^{p} \alpha_k z^{-k}\right]. \tag{A.27}$$

Using the LP filter coefficients $\{\alpha_k\}_{k=1,...,p}$, and expanding the previous equation yields the following solution to determining the cepstral parameters:

$$c_1 = -\alpha_1 \tag{A.28}$$

$$c_j = -\alpha_j - \sum_{k=1}^{j-1}(\frac{k}{j})c_k\alpha_{j-k}, \quad j = 2,3,...,p, \tag{A.29}$$

$$c_j = -\sum_{k=1}^{p}[(\frac{j-k}{j})c_{j-k}]a_k, \quad j = p+1, p+2... \tag{A.30}$$

In our system, $p = 10$ and we truncate the cepstral coefficients at c_{40}, or in other words $j = 11, 12, ..., 40$ for equation (A.30).

We will now define the distance measures and see how these parameters we have derived are used in their calculations.

A.2 Calculating the Average Speech Distortion Measure

A.2.1 A Cosh Measure

The average speech distortion contains four different *cosh* distance measures. Their difference will be discussed later, but the cosh measure's general form will be next revealed.

It can be shown that equation (A.9) can be rewritten as follows [12]

$$E = \sum_{k=-p}^{p} r_a(k) r_x(k) \tag{A.31}$$

$$= r_a(0) r_x(0) + 2 \sum_{k=1}^{p} r_a(k) r_x(k), \tag{A.32}$$

where equation (A.32) follows from the symmetry of autocorrelation functions ($r(-k) = r(k)$). Note, the value of E is also called the *minimum residual error* since it has been minimized in selecting the LP inverse filter coefficients. $E^{'}$ can be defined similarly to E where $r^{'}_a(k)$ and $r^{'}_x(k)$ are used instead.

The *residual energy*, δ, is obtained when a test speech sequence is passed through a filter for which it has not been optimized. In other words, if we pass the input speech sequence, $\{x(m)\}$, through the output speech filter, $A^{'}(z)$, we obtain the input residual energy [12],

$$\delta \triangleq r^{'}_a(0) r_x(0) + 2 \sum_{k=1}^{p} r^{'}_a(k) r_x(k). \tag{A.33}$$

The output residual energy, $\delta^{'}$, can be thought of as the opposite process.

This leads us to the definition of two *likelihood ratios*, δ/E and $\delta^{'}/E^{'}$, that define the difference between the test and reference speech sequences. Note, that these likelihood measures are always greater than 1, with equality only if the two filters, $A(z)$ and $A^{'}(z)$, are equal.

The cosh measure, Ω, can now be defined as [12]

$$\Omega \triangleq \frac{1}{2}\left(\frac{\sigma}{\sigma^{'}}\right)^2\left(\frac{\delta}{E}\right) + \frac{1}{2}\left(\frac{\sigma^{'}}{\sigma}\right)^2\left(\frac{\delta^{'}}{E^{'}}\right) - 1, \tag{A.34}$$

where σ and $\sigma^{'}$ are the respective gains of the input and output LP filters. We now define ω so that we can express the cosh measure in terms of decibels, in other words,

$$\Omega = cosh(\omega) - 1; \tag{A.35}$$

therefore,

$$\omega = ln[1 + \Omega + \sqrt{\Omega(2+\Omega)}]. \tag{A.36}$$

A.2.2 A Cepstral Measure

The last three distance measures in the average speech distortion are various *cepstral distance* measures, which we will now define. First, we note some properties of the cepstral coefficients $\{c_k\}_{k=1,\ldots,p}$:

$$\begin{aligned} c_0 &\triangleq ln[\sigma^2], & \text{(A.37)} \\ c_{-k} &= c_k. & \text{(A.38)} \\ & & \text{(A.39)} \end{aligned}$$

Now we can define cepstral distance between two sets of coefficients, $\{c_k\}$ and $\{c_k^{'}\}$, as

$$[u(p)]^2 \triangleq \sum_{k=-p}^{p}(c_k - c_k^{'})^2 \tag{A.40}$$

$$= \quad (c_0 - c_0^{'})^2 + 2\sum_{k=1}^{p}(c_k - c_k^{'})^2. \qquad \text{(A.41)}$$

A.2.3 Choice of LP Filter Gain

The previously mentioned difference in the cosh or cepstral measures lies in the choice of gain. We now outline four possible choices for σ and $\sigma^{'}$ [12]:

- The first is to set both gains to unity, ie. $\sigma = \sigma^{'} = 1$.
- A second is to define the spectral energies to be equal. Thus, $\sigma^2 = E/r_x(0)$ and $(\sigma^{'})^2 = E^{'}/r_x^{'}(0)$.
- The third choice is to choice the gains so as to make the model spectral energy equal to the energy of the original signal. In other words set $\sigma^2 = E$ and $(\sigma^{'})^2 = E^{'}$.
- The final choice applies only to cosh measures and entails choosing the gains to minimize the cosh measure. The minimized value of the cosh measure is found to be

  $$\Omega_{min} = \sqrt{\left(\frac{\delta}{E}\right)\left(\frac{\delta'}{E'}\right)} - 1, \qquad \text{(A.42)}$$

  where, as before,

  $$\omega_{min} = ln[1 + \Omega_{min} + \sqrt{\Omega_{min}(2 + \Omega_{min})}]. \qquad \text{(A.43)}$$

A.2.4 Average Speech Distortion

The average speech distortion is simply an equal average of all the distance measures listed below:

$$D_1 = \quad 10log_{10}e \cdot \omega \qquad \text{for } \sigma = \sigma^{'} = 1, \qquad \text{(A.44)}$$

$$
\begin{aligned}
D_2 = &\quad 10log_{10}e \cdot \omega \qquad \text{for } \sigma^2 = E/r_x(0),\ (\sigma^{'})^2 = E^{'}/r_x^{'}(0), && \text{(A.45)}\\
D_3 = &\quad 10log_{10}e \cdot \omega \qquad \text{for } \sigma^2 = E,\ (\sigma^{'})^2 = E^{'}, && \text{(A.46)}\\
D_4 = &\quad 10log_{10}e \cdot \omega_{min} && \text{(A.47)}\\
D_5 = &\quad 10log_{10}e \cdot u[p] \qquad \text{for } \sigma = \sigma^{'} = 1, && \text{(A.48)}\\
D_6 = &\quad 10log_{10}e \cdot u[p] \qquad \text{for } \sigma^2 = E/r_x(0),\ (\sigma^{'})^2 = E^{'}/r_x^{'}(0), && \text{(A.49)}\\
D_7 = &\quad 10log_{10}e \cdot u[p] \qquad \text{for } \sigma^2 = E,\ (\sigma^{'})^2 = E^{'}, && \text{(A.50)}
\end{aligned}
$$

where, ω is given by equation (A.36), ω_{min} is given by equation (A.43), and $u[p]$ is given by equation (A.41). Note, that the $10log_{10}e$ factor converts all the distance measures into decibels.

Therefore, the speech distortion measure for a single subframe at time n, is given by

$$D(n) = \frac{1}{7}\sum_{i=1}^{7} D_i, \tag{A.51}$$

which is then average over all the subframes to give

$$D_{avg} = \frac{1}{N}\sum_{n=1}^{N} D(n), \tag{A.52}$$

where N is the total number of subframes.

Bibliography

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

[2] F. I. Alajaji, N. Phamdo, N. Farvardin, and T.E. Fuja, "Detection of Binary Markov Sources over Channels with Additive Markov Noise," *IEEE Transactions on Information Theory*, vol. IT-42, pp. 230-38, Jan. 1996.

[3] F. I. Alajaji, S. Al-Semari and P. Burlina, "An Unequal Error Protection Trellis Coding Scheme for Still Image Communication," *Proc. International Symposium on Information Theory and its Applications*, Ulm, Germany, June 1997.

[4] S. Al-Semari, F. I. Alajaji, and T. E. Fuja, "Sequence MAP Decoding of Trellis Codes for Gaussian and Rayleigh Channels," *Proc. International Symposium on Information Theory and its Applications*, Victoria, Sept. 1996.

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

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

[7] T. M. Cover, and J. A Thomas, *Elements of Information Theory*, New York: John Wiley & Sons, Inc., 1991.

[8] J. R. Deller, Jr., J. G. Proakis, J.H.L. Hansen, *Discrete-Time Processing of Speech Signals*, New York: Macmillan Publishing Company, 1993.

[9] N. Farvardin, V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An approach to combined source-channel coding," *IEEE Transactions on Information Theory*, vol. IT-33, pp. 827-838, Nov. 1987.

[10] N. Farvardin, and R. Laroia, "Efficient Encoding of Speech LSP Parameters Using the Discrete Cosine Transformation," *Proc. ICASSP 1989*, pp.168-71.

[11] G.D. Forney, Jr., "Convolutional Codes II: Algebraic Structure," *Information and Control*, Vol. 25, pp.222-266, July 1974.

[12] A.H. Gray, and J.D. Markel, "Distance Measures for Speech Processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-24, no. 8, Oct. 1976.

[13] D. Haccoun, and G. Begin, "High-rate Punctured Convolutional Codes for Viterbi and Sequential Decoding," *IEEE Transactions on Communications*, vol. 37, no. 11, pp. 1113-25, November 1989.

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

[15] J. Hagenauer, "Joint Source-Channel Coding for Broadcast Applications," *Audio and Video Digital Radio Broadcasting Systems and Techniques*. New York: Elsevier, 1994.

[16] J. Hagenauer, N. Seshadri, and C. W. Sundberg, "The Performance of Rate-Compatible Punctured Convolutional Codes for Digital Mobile Radio," *IEEE Transactions on Communications*, vol. 38, no. 7, pp. 966-980, July 1990.

[17] F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-23, no. 1, Feb. 1975.

[18] J. M. Kroll and N. Phamdo, "Optimization of Four Dimensional 8-PSK Trellis Codes for a Binary Markov SOurce with MAP Detection," *Proceeding of Canadian Workshop on Information Theory*, Toronto, June 1997.

[19] R. Laroia, N. Phamdo, and N. Farvardin, "Robust and Efficient Quantization of Speech LSP Parameters Using Structured Vector Quantization," *Proc. ICASSP 1991*, pp. 661-664.

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

[21] L.H.C. Lee, *Convolutional Coding: Fundamentals and Applications*, Norwood: Artech House, Inc., 1997.

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

[23] J.W. Modestino, and D.G. Daut, "Combined Source-Channel Coding of Images," *IEEE Transactions on Communications*, vol. 27, pp.1644-59, November 1979.

[24] J.W. Modestino, D.G. Daut, and A.L.Vickers, "Combined Source-Channel Coding of Images Using the Block Cosine Transform," *IEEE Transactions on Communications*, vol. 29, pp.1261-74, September 1981.

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

[26] National Communications System (NCS) "Telecommunications: Analog to Digital Conversion of Radio Voice by 4800 Bit/Second Code Excited Linear Prediction (CELP)," Office of Information Resource Management, NCS, Arlington, VA, Federal Standard 1016.

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

[28] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3^{rd} Ed., New York: McGraw-Hill, Inc., 1991.

[29] L.R. Rabiner, and R.W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, New Jersey: Prentice Hall, 1978.

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

[31] D.E. Ray, and D.J. Rahikka, "Reed-Solomon Coding for CELP EDAC in Land Mobile Radio," *Proceedings of ICASSP 1994*, Adelaide, Australia, April 19-22, 1994.

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

[33] W. Xu, J. Hagenauer, and J. Hollmann, "Joint Source-Channel Decoding Using Residual Redundancy in Compressed Images," *Proc. International Conference on Communications*, Dallas, 1996.

Vita

Ali Nazer

Education:

1995 to 1997	M.Sc. (Eng.)	Queen's University	Mathematics and Engineering
1991 to 1995	B.Sc. (Honours 1^{st} Class)	Queen's University	Electrical Engineering

Experience:

1995 to 1997	Research Assistant, Queen's University
1996 to 1997	Teaching Assistant, Queen's University
1995	Volunteer, Queen's Project on International Development
1994 to 1995	Power Control Systems Designer, Queen's Solar Vehicle Team