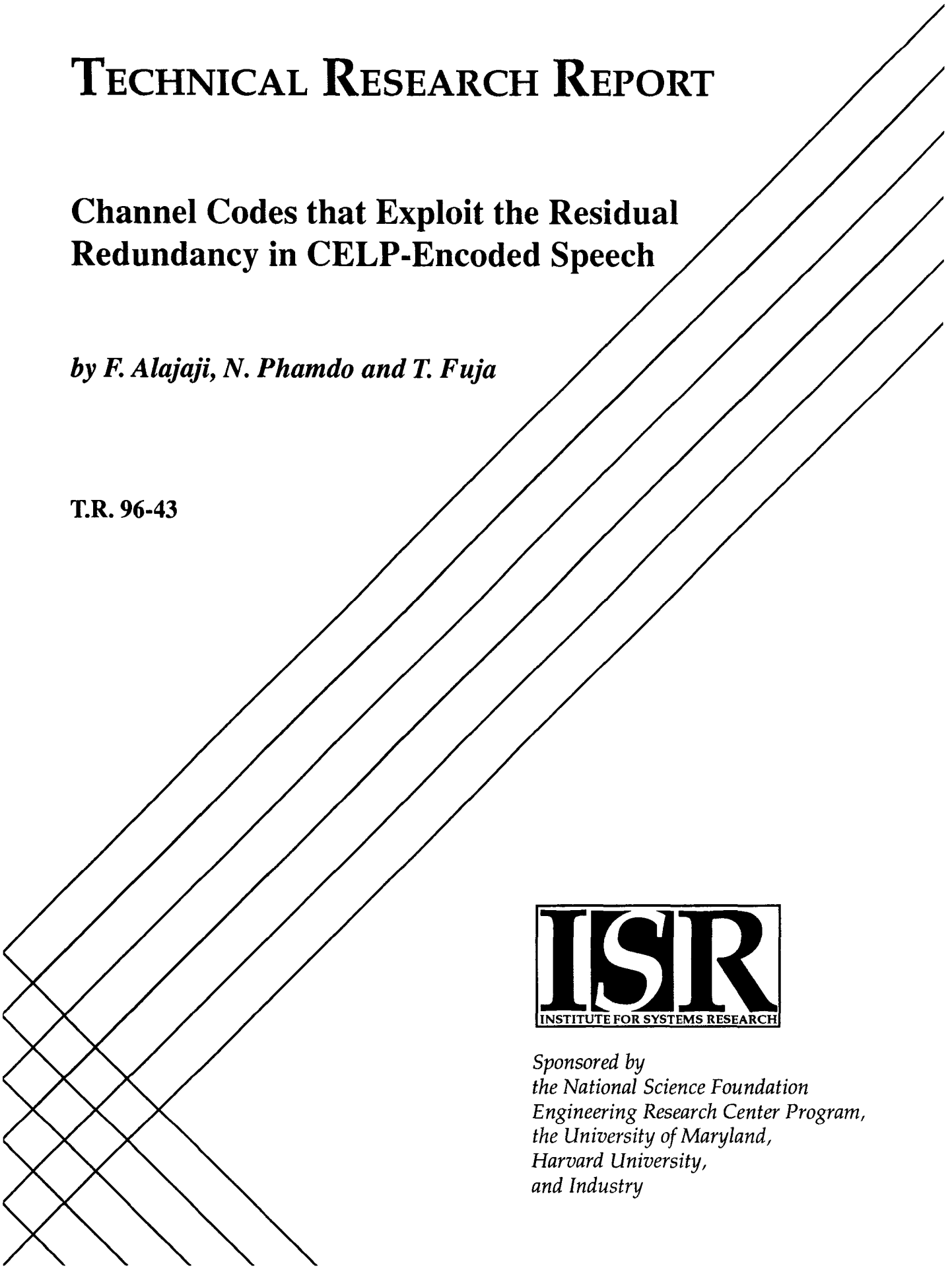
TECHNICAL RESEARCH REPORT

Channel Codes that Exploit the Residual Redundancy in CELP-Encoded Speech

by F. Alajaji, N. Phamdo and T. Fuja

T.R. 96-43



Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Channel Codes That Exploit the Residual Redundancy in CELP-Encoded Speech*

Fady Alajaji†, *Nam Phamdo*‡ and *Tom Fuja**

† Dept. of Mathematics and Statistics
Queens University
Kingston, Ontario K7L 3N
Canada

‡ Electrical Engineering Dept.
State University of New York
at Stony Brook
Stony Brook, NY 11794-2350

* Electrical Engineering Dept.
Institute for Systems Research
University of Maryland
College Park, MD 20742

To Appear in Sept. 1996 IEEE Transactions on Speech and Audio Processing

Abstract

We consider the problem of reliably transmitting CELP-encoded speech over noisy communication channels. Our objective is to design efficient coding/decoding schemes for the transmission of the CELP line spectral parameters (LSP's) over very noisy channels.

We begin by quantifying the amount of "residual redundancy" inherent in the LSP's of Federal Standard 1016 CELP. This is done by modeling the LSP's as first- and second-order Markov chains. Two models for LSP generation are proposed; the first model characterizes the intra-frame correlation exhibited by the LSP's, while the second model captures both intra-frame and inter-frame correlation. By comparing the entropy rates of the models thus constructed with the CELP rates, it is shown that as many as one-third of the LSP bits in every frame of speech are redundant.

We next consider methods by which this residual redundancy can be exploited by an appropriately designed channel decoder. Before transmission, the LSP's are encoded with a forward error control (FEC) code; we consider both block (Reed-Solomon) codes and convolutional codes. Soft-decision decoders that exploit the residual redundancy in the LSP's are implemented assuming additive white Gaussian noise (AWGN) and independent Rayleigh fading environments. Simulation results employing binary phase-shift keying (BPSK) indicate coding gains of 2 to 5 dB over soft-decision decoders that do *not* exploit the residual redundancy.

EDICS Category: SA 1.4.8 – Combined Source and Channel Coding.

*The work of Alajaji (fady@polya.mast.queensu.ca) and Fuja (fuja@eng.umd.edu) was supported in part by the Department of Defense; also by NSF grant NCR-8957623 and by the NSF Engineering Research Centers Program, CDR-8803012. The work of Phamdo (phamdo@sbee.sunysb.edu) was supported in part by NTT Corporation.

I. Introduction

The "gospel according to Shannon" tells the communication system designer that a system's source code and its channel code should be designed independently of one another. Shannon's separation principle [1] mandates that whatever performance is achievable with a jointly-designed source-code/channel-code is also achievable with a source code designed solely with regard to the source description and a channel code designed solely with regard to the channel description.

Of course, the separation principle is an *asymptotic* result – one that permits unlimited complexity and delay in the encoding/decoding operations. It is possible that – for a fixed degree of delay and/or complexity – the best jointly-designed source-code/channel-code could out-perform the best separately-designed pair. This observation has led a number of researchers to consider schemes in which the source code and channel code are designed jointly [2]-[5].

Most of the work on joint source-channel coding has focused on the design of source codes that are robust in the face of channel errors. By contrast, Hagenauer [5] has looked at methods by which the overall system performance can be improved by designing the channel decoder to exploit the known characteristics of the source code – an approach he calls "source controlled channel decoding." The work in this paper is similar in spirit to [5].

We consider the problem of reliably transmitting speech encoded using codebook-excited linear predictive (CELP) coding over a noisy communication channel. CELP coding is a frame-oriented technique that breaks a speech signal into blocks of samples that are processed as one unit. The particular implementation we consider is Federal Standard 1016 (FS 1016) 4.8 kbit/s CELP [6]. The CELP parameters that are transmitted over the noisy channel include the stochastic code book index and gain, the adaptive code book index (pitch delay) and gain, and 10 ordered line spectral parameters (LSP's).

An "ideal" source encoder would accept the signal to be compressed and produce an

independent, identically distributed (i.i.d.) sequence of equiprobable bits at the output. If the source encoder output is *not* i.i.d. and equiprobable, then it is usually possible to compress it even further.

Most source encoders are not ideal; certainly, CELP is not. As a result the bitstream produced by a CELP encoder is *not* i.i.d. equiprobable, and so the number of bits produced per unit time is significantly greater than the entropy rate of the output. This *residual redundancy* reflects the *residual correlation* and the *non-uniformity* of the encoded bitstream; the CELP encoder leaves some redundancy in the encoded bitstream in the form of memory and non-uniformity.

In this paper, we consider methods by which channel codes can take advantage of this residual redundancy to enhance the performance of CELP-encoded speech over very noisy channels. Specifically, we investigate techniques by which the residual redundancy inherent in the line spectral parameters (LSP's) of CELP-encoded speech can be quantified and exploited. We begin by proposing two models for the generation of LSP's. The first model incorporates only the non-uniformity of the LSP's and their correlation within a CELP frame; the second model provides for correlation between frames as well. When these models are "trained" using an actual CELP bitstream they show that as many as 12.5 of the 30 high-order LSP bits in each frame may be redundant.

Once the residual redundancy in the LSP's is quantified, we present decoding algorithms that exploit that redundancy via both convolutional and block (Reed-Solomon) codes. In the case of convolutional codes, we employ three optimal soft-decision decoding schemes, all based on the Viterbi algorithm:

- ML – the "usual" maximum likelihood (ML) decoding algorithm.
- MAP 1 – a maximum a-posteriori (MAP) decoding algorithm that exploits only the redundancy due to the non-uniform distribution of the LSP's and their correlation within a frame – approximately 10 bits/frame.

- MAP 2 – which exploits the redundancy from the non-uniform distribution of the LSP's *and* their inter-frame and intra-frame correlation – approximately 12.5 bits/frame.

All three algorithms are implemented so as to yield a decoding delay of only one frame.

In the case of block coding, we present four low complexity sub-optimal soft-decision decoding (SDD) algorithms; the first three schemes are codeword-by-codeword decoding algorithms, while the last one is sequence-based:

- SDD 1 – which approximates "traditional" maximum likelihood decoding and does not attempt to exploit any of the residual redundancy.

- SDD 2 – which exploits only the redundancy due to the ordered nature of the LSP's – approximately 4.4 bits/frame.

- SDD 3 – which like MAP 1 exploits the redundancy due to the non-uniform distribution of the LSP's and their correlation within a frame – around 10 bits/frame of redundancy.

- SDD 4 – which like MAP 2 exploits both the inter- and intra-frame correlation and the redundancy due to the non-uniform distribution – approximately 12.5 bits/frame of redundancy.

Simulation results show that the decoding algorithms that exploit the most residual redundancy obtain the most coding gain – especially over very noisy channels.

The rest of this paper is organized as follows. In Section II, two models for LSP generation are proposed and the resulting estimates of the residual redundancy are presented. The channel models are briefly described in Section III. The convolutional and Reed-Solomon codes as well as their soft-decision decoding algorithms are investigated in Sections IV and V, respectively. In Section VI, simulation results on the performance of the decoders are presented and analyzed. Results of listening tests for subjective evaluation are introduced in Section VII. Finally, conclusions are stated in Section VIII.

II. LSP Residual Redundancy

In this section, we quantify the residual redundancy in the encoded LSP's of FS 1016 CELP [6]. In FS 1016 CELP, each LSP is quantized by either a three-bit or a four-bit scalar quantizer. The second through fifth LSP's are quantized by four-bit quantizers; the rest are quantized to three bits. The quantized LSP's are guaranteed to be ordered (LSP-1 < LSP-2 $< \cdots <$ LSP-10). In this paper, we consider only the three most significant bits of each LSP, ignoring the least significant bit in the second through fifth parameters.

Suppose we encode a segment of speech using FS 1016 CELP, resulting in a sequence of CELP frames. Let $\{U_{i,j} : 1 \leq i \leq 10, j = 1, 2, \ldots\}$ denote a random process in which $U_{i,j}$ is the i^{th} (three-bit) quantized LSP in frame j. Let $\mathbf{U}_j = [U_{1,j}, U_{2,j}, \ldots, U_{10,j}]$ denote the vector consisting of the 10 quantized LSP's in frame j. If we assume that this random process is stationary[1] then the entropy rate (in bits/frame) of this process is given by

$$H_{\mathrm{F}} = \lim_{n \to \infty} \frac{1}{n} H(\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n),$$

where

$$H(\mathbf{U}_1, \ldots, \mathbf{U}_n) = - \sum_{\mathbf{u}_1, \ldots, \mathbf{u}_n} \Pr(\mathbf{U}_1 = \mathbf{u}_1, \ldots, \mathbf{U}_n = \mathbf{u}_n) \log_2 \Pr(\mathbf{U}_1 = \mathbf{u}_1, \ldots, \mathbf{U}_n = \mathbf{u}_n).$$

H_{F} represents the minimum number of bits per frame required to describe $\{U_{i,j} : 1 \leq i \leq 10, j = 1, 2, \ldots\}$. If we assume $U_{i,j}$ is represented by three bits, then the CELP encoder produces 30 bits/frame to describe the LSP's, so the *residual redundancy* – i.e., the total redundancy (per frame) in the CELP-encoded LSP's – is

$$\rho_T = 30 - H_{\mathrm{F}} \text{ (bits/frame)}.$$

We seek to estimate H_{F} (and so ρ_T). We do this by observing a long training sequence – i.e., a realization of $\{U_{i,j} : 1 \leq i \leq 10, j = 1, 2, \ldots\}$ – and matching the observations to

[1] We do not claim that the LSP's form a stationary random process. However, to the extent that the quantized LSP's can be *approximated* by a stationary random process, the following calculations indicate how much redundancy is inherent in those quantized LSP's.

a particular model of a random process; we then compute the entropy rate of the model process and use that as our estimate of H_{F}.

The two models are as follows:

- Model A assumes that the LSP's in two different frames are independent, and the LSP's within a frame form a first-order Markov chain. More specifically, it assumes that

$$\Pr(\mathbf{U}_j = \mathbf{u}_j | \mathbf{U}_1 = \mathbf{u}_1, \mathbf{U}_2 = \mathbf{u}_2, \ldots, \mathbf{U}_{j-1} = \mathbf{u}_{j-1}) = \Pr(\mathbf{U}_j = \mathbf{u}_j),$$

and

$$\begin{aligned} &\Pr(U_{i,j} = u_{i,j} | U_{1,j} = u_{1,j}, \ldots, U_{i-1,j} = u_{i-1,j}) \\ &\qquad = \Pr(U_{i,j} = u_{i,j} | U_{i-1,j} = u_{i-1,j}), \\ &\qquad = P_A^{(i)}(u_{i,j} | u_{i-1,j}), \end{aligned} \tag{1}$$

for $i = 2, 3, \ldots, 10$ and $j = 1, 2, \ldots$. For $i = 1$, the above becomes $P_A^{(1)}(u_{1,j})$.

- Model B is more complex; it assumes a second-order Markov structure – that $U_{i,j}$ is independent of all the LSP's that came before it conditioned on knowing $U_{i-1,j}$ and $U_{i,j-1}$ – the LSP that immediately precedes it in the same frame and the corresponding LSP in the frame immediately preceding. More specifically, it assumes

$$\begin{aligned} &\Pr(U_{i,j} = u_{i,j} | \mathbf{U}_1 = \mathbf{u}_1, \ldots, \mathbf{U}_{j-1} = \mathbf{u}_{j-1}, U_{1,j} = u_{1,j} \ldots, U_{i-1,j} = u_{i-1,j}) \\ &\qquad = \Pr(U_{i,j} = u_{i,j} | U_{i,j-1} = u_{i,j-1}, U_{i-1,j} = u_{i-1,j}), \\ &\qquad = P_B^{(i)}(u_{i,j} | u_{i,j-1}, u_{i-1,j}), \end{aligned} \tag{2}$$

$i = 2, 3, \ldots, 10$, $j = 2, 3, \ldots$. For $i = 1$, the above becomes $P_B^{(1)}(u_{1,j} | u_{1,j-1})$.

Note that in (1) and (2), the probability transition matrix depends on i but not on j.

Remark: The assumption of a Markov structure both temporally (i.e., between frames) and within a frame is based on two observations:

1. Error masking is performed on LSP's by interpolating between the corresponding LSP's in adjacent frames; the success of this approach suggests the temporal Markov structure.

2. The ordering property of the LSP's within a frame suggest an intra-frame Markov structure.

We do not attempt to justify these models more rigorously. The extent to which a model "fits" will be judged on the basis of how well decoding algorithms matched to that model perform.

Procedure: A large training sequence from the TIMIT speech database [13] was used; for every 30 msec of speech an LPC analysis was performed according to FS 1016 standards to arrive at the 10 quantized LSP's. The relative frequency of transitions between the values of the three high-order bits of each LSP were compiled to extract Markov transition probabilities for Model A and Model B. The entropy of the resulting Markov chains was computed to arrive at an estimate of the redundancy in each LSP and in each frame. Let $H^* = \sum_{i=1}^{10} H(U_{i,j})$ be the process entropy rate (in bits/frame) *if the LSP's were independent* (H^* is independent of j since (1) and (2) are independent of j). Note that we can write $\rho_T = \rho_D + \rho_M$ where $\rho_D \stackrel{\Delta}{=} 30 - H^*$ denotes the frame redundancy due to the non-uniform distribution of the LSP's and $\rho_M \stackrel{\Delta}{=} H^* - H_{\mathrm{F}}$ denotes the frame redundancy due to the memory between the LSP's [7].

The results are compiled in Table 1 in which we provide the values of ρ_D, ρ_M and ρ_T for each individual LSP as well as for the entire frame.

- Model A – which does not attempt to take into account *any* correlation between frames – indicates that $\rho_T = 9.867$ of the 30 high-order bits in the LSP's are redundant. Approximately $\rho_D = 5.275$ bits of redundancy were due to the non-uniform distribution of the LSP's, and approximately $\rho_M = 4.593$ bits of redundancy were due to the memory within a frame.

- Model B – which *does* take into account both inter-frame and intra-frame correlation – indicates that $\rho_T = 12.485$ of the 30 high-order bits in the LSP's are redundant. Once again, $\rho_D = 5.275$ bits of redundancy were due to the non-uniform distribution of the LSP's, while $\rho_M = 7.211$ bits of redundancy were due to the memory remaining both within a frame and between frames.

Clearly, substantial redundancy exists within the LSP's. Using Model B as our guide, there are only 17.515 bits of "real" information buried in the 30 high-order LSP bits; it is as though the CELP encoder had passed the description of the LSP's through a (poorly designed) rate $17.5/30 \approx 0.584$ channel encoder. In the Sections IV and V we discuss methods for exploiting that redundancy.

Finally, the amount of redundancy due to the LSP ordering property can be computed. It can be shown using the counting algorithm of [12] that, of the 2^{30} possible high-order bits in the quantized LSP's, only $50,644,887$ of them correspond to ordered LSP's. Thus the redundancy due to ordering is $30 - \log_2[5.06 \cdot 10^7] = 4.406$ bits. This redundancy is embedded in the total redundancy shown in Table 1.

III. Channel Models

The channels considered in this paper are the additive white Gaussian noise (AWGN) channel and the fully interleaved Rayleigh fading channel, both used with BPSK modulation. More specifically, we assume the j^{th} received signal y_j is related to the j^{th} transmitted signal x_j according to

$$y_j = a_j x_j + n_j.$$

Here, $x_j \in \{+\sqrt{E_s}, -\sqrt{E_s}\}$ and n_j is a zero-mean Gaussian random variable with variance $N_0/2$. (We assume n_i and n_j are independent for $i \neq j$.) The distribution of the fading coefficient a_i depends on the channel assumption:

- For a purely AWGN channel, we assume $a_i = 1$.

- For the fully interleaved Rayleigh fading channel we assume a_i has a Rayleigh distribution with $E[a_j^2] = 1$. Moreover, to accommodate the "fully interleaved" description we assume that a_i and a_j are independent for $i \neq j$.

IV. Decoding Schemes for a Convolutionally Encoded System

In this section we propose soft-decision decoding algorithms for a convolutionally encoded system.

The (three-bit) quantized LSP's are channel encoded with the 32-state rate-3/4 convolutional encoder in [11, p. 331]. This code has minimum free Hamming distance $d_{\text{free}} = 5$ and generator matrix

$$G(D) = \begin{pmatrix} 1+D & D & D & 1+D \\ D^2 & 1+D & 0 & 1+D+D^2 \\ 0 & D & 1+D^2 & 1+D^2 \end{pmatrix}.$$

The input to the encoder is a sequence of three-bit LSP's, resulting in a sequence of four-bit outputs; for each frame, LSP-1 is encoded first, followed by LSP-2, LSP-3, etc. Once LSP-10 is encoded, LSP-1 from the *next* frame is encoded.

Let $u_{i,j}$ denote the i^{th} quantized LSP from frame j, where $1 \leq i \leq 10$ and $j \in \{1, 2, \ldots\}$. Letting $k = 10(j-1) + i$, re-index $u_{i,j}$ as u_k, so $[u_1, u_2, \ldots u_{10}]$ are the quantized LSP's from the first CELP frame, $[u_{11}, u_{12}, \ldots, u_{20}]$ are the quantized LSP's from the second CELP frame, etc. Let $\mathbf{x}_k \in \{+\sqrt{E_s}, -\sqrt{E_s}\}^4$ be the real 4-tuple generated by the BPSK modulator in response to $\mathbf{u}_k$. Similarly, let $\mathbf{a}_k$ and $\mathbf{y}_k$ be the real 4-tuples corresponding to the associated fading coefficients and channel outputs – i.e.,

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

where $\mathbf{a}_k \mathbf{x}_k$ is the component-wise product of $\mathbf{a}_k$ and $\mathbf{x}_k$, and $\mathbf{n}_k$ is a 4-tuple of independent, zero-mean Gaussian random variables, each with variance $N_0/2$. (If the channel is Rayleigh fading; then $\mathbf{a}_k$ consists of four independent Rayleigh random variables; if the channel is AWGN, then $\mathbf{a}_k$ is the all-one vector.)

We consider three soft-decision decoding schemes based on the Viterbi algorithm.

- **ML:** This decoding algorithm is maximum likelihood decoding. The decoder chooses the code sequence $\{\mathbf{x}_k\}$ that minimizes

$$\sum_{k=1}^{K} ||\mathbf{y}_k - \mathbf{a}_k\mathbf{x}_k||^2,$$

where K is the number of received symbols (a multiple of 10).

- **MAP 1:** This decoder is a maximum a posteriori decoder that exploits the source redundancy inherent in Model A. Here, the decoder chooses the code sequence $\{\mathbf{x}_k\}$ that minimizes

$$\sum_{k=1}^{K} ||\mathbf{y}_k - \mathbf{a}_k\mathbf{x}_k||^2 - N_0 \ln P_A^{([k \bmod 10])}(u_k|u_{k-1}),$$

where $\{u_k\}$ is the LSP sequence corresponding to the code sequence $\{\mathbf{x}_k\}$, and $[k \bmod 10]$ is the unique integer between 1 and 10 that is equivalent to k modulo 10. [2] Note that if $k+9$ is a multiple of 10 – i.e., $k \in \{1, 11, 21 \ldots\}$ – then $P_A^{([k \bmod 10])}(u_k|u_{k-1}) = P_A^{(1)}(u_k)$, the probability of the first LSP in frame $(k+9)/10$.

- **MAP 2:** This is similar to the above except that the decoder exploits the source redundancy inherent in Model B. The goal here is to minimize

$$\sum_{k=1}^{K} ||\mathbf{y}_k - \mathbf{a}_k\mathbf{x}_k||^2 - N_0 \ln P_B^{([k \bmod 10])}(u_k|u_{k-10}, u_{k-1}).$$

In the simulation, the decoder has a path memory of 10 (five times the encoder memory [11, p. 338]). That is, upon receiving $\mathbf{y}_k$, the decoder releases the estimate $\hat{u}_{k-10}$. The estimate, $\hat{u}_{k-10}$, is chosen as the symbol on the path that has minimum cumulative distortion up to time-index k. This symbol release rule corresponds to a decoding delay of one frame.

In the above algorithms, if the decoded LSP vector $\hat{U}_j$ is not ordered, we simply re-order them to yield an ordered output. For the two MAP decoders, the probability of an unordered decoded LSP vector is small but non-zero because we have ignored the least significant bit in LSP's 2 through 5. These four least significant bits are also BPSK-modulated and sent over

[2] Our definition of $[k \bmod 10]$ differs slightly from the "usual" one which defines it as an integer between 0 and 9. We define $[k \bmod 10]$ as we do because the LSP's are numbered 1-10 instead of 0-9.

the noisy channel. They are, however, not protected by the convolutional code and therefore may (with higher probability) be corrupted.

V. Decoding Schemes for the Reed-Solomon Encoded System

We now describe four different soft-decision decoding (SDD) algorithms for block codes. In each case we assume that the ten three-bit LSP's in each frame are encoded using a $(15, 10)$ code $\mathcal{C}$ over $\mathcal{F}_8$ (the Galois field with eight elements). The particular code we use is the direct sum of a $(9, 6)$ extended Reed-Solomon code with $d_{\min} = 4$ (call this code $\mathcal{C}_1$) and a $(6, 4)$ shortened Reed-Solomon code with $d_{\min} = 3$ (call this code $\mathcal{C}_2$). The resulting direct sum has a generator matrix

$$G = \begin{pmatrix} G_1 & 0 \\ 0 & G_2 \end{pmatrix},$$

where

$$G_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \alpha^4 & \alpha^5 & \alpha^2 \\ 0 & 1 & 0 & 0 & 0 & 0 & \alpha & \alpha^3 & \alpha^4 \\ 0 & 0 & 1 & 0 & 0 & 0 & \alpha^4 & 1 & \alpha^6 \\ 0 & 0 & 0 & 1 & 0 & 0 & \alpha^2 & \alpha^6 & \alpha \\ 0 & 0 & 0 & 0 & 1 & 0 & \alpha^2 & 1 & \alpha^3 \\ 0 & 0 & 0 & 0 & 0 & 1 & \alpha & 1 & \alpha^5 \end{pmatrix}$$

and

$$G_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & \alpha^3 & \alpha^4 \\ 0 & 1 & 0 & 0 & \alpha^3 & \alpha^5 \\ 0 & 0 & 1 & 0 & \alpha & \alpha^5 \\ 0 & 0 & 0 & 1 & \alpha & \alpha^4 \end{pmatrix},$$

where α is a primitive root of $x^4 + x + 1$ [11]. $\mathcal{C}_1$ protects the first six LSP's while $\mathcal{C}_2$ protects the last four LSP's.

We then assume that the 15-symbol codewords are transmitted over either the BPSK-modulated AWGN or Rayleigh channels. The four soft-decision decoding algorithms make increasing use of the residual redundancy present in the LSP's. In each algorithm, the two codes are decoded sequentially. First, $\mathcal{C}_1$ is decoded to obtain the first six LSP's and then $\mathcal{C}_2$ is decoded to obtain the last four LSP's. Descriptions of the decoding algorithms follow.

- **SDD 1:** This decoding algorithm is near-maximum likelihood (near-ML) codeword-by-codeword decoding. We break the decoding into Stage 1 and Stage 2.

- Stage 1 generates a list of candidates that will "compete" to be the decoder's estimate of the transmitted codeword. The matched filter outputs corresponding to the encoded LSP's are quantized into 15-tuples over $\mathcal{F}_8$ in $P = 2^b$ different ways; this is done by toggling the b least confident bits. Then for each of the P quantizations we arrive at L different estimates of the noise, corresponding to the L "lightest" vectors in the same coset as the quantized 15-tuple; the net result is at most $N = PL$ codeword candidates. We say "at most" because the candidates thus generated need not be distinct. To generate a large list while simultaneously bounding the required complexity, we propose quantizing the matched filter outputs until there are N distinct codeword candidates *or* until we have quantized the matched filter outputs $P_{\max}$ times – whichever comes first.
- Stage 2 compares the distance between the received (unquantized) vector and each of the codeword candidates on the list. More specifically, if $\mathbf{y} = [y_0, y_1, \ldots, y_{44}]$ is the real 45-tuple corresponding to the (unquantized) matched filter outputs, then we choose as our estimate the codeword

$$\mathbf{c}^* = \arg\min \left\{ \sum_{j=0}^{44} (y_j - a_j x_j)^2 : \mathbf{c} = [c_0, c_1, \ldots, c_{44}] \text{ is on list.} \right\}, \qquad (3)$$

  where $\mathbf{x} = [x_0, x_1, \ldots, x_{44}] \in \mathbf{R}^{45}$ denote the BPSK-modulated codeword $\mathbf{c} = [c_0, c_1, \ldots, c_{44}]$ and a_j is the fading coefficient. Note that with this algorithm, the decoded LSP vector might not be ordered. In such a case, the LSP vector is re-ordered to yield an ordered output vector [9].

If $N = P = P_{\max} = 1$, then SDD 1 reduces to hard-decision maximum likelihood (HDML) decoding. Furthermore, if $L = 1$ – so that a single "best estimate" is generated for each quantization of the matched filter outputs – then SDD 1 is essentially equivalent to a non-binary Chase algorithm [8] in which $N = P = 2^b$ different error "test patterns" are considered.

- **SDD 2:** This algorithm is identical to SDD 1 with one exception: The ordering of the LSP's is taken into account. More specifically, during Stage 1 we now generate a list of at most N codewords corresponding to *ordered* LSP's. This is done by quantizing the matched filter outputs until N such distinct codewords have been generated or until P_{max} quantizations have been performed – whichever happens first. Stage 2 consists, once again, of choosing from the list the codeword that minimizes the weighted Euclidean distance to the (unquantized) received vector. If the list is empty, we simply repeat the LSP's from the previous frame.

- **SDD 3:** This algorithm is near-maximum a posteriori (near-MAP) codeword-by-codeword decoding. It is identical to SDD 2 except that, during Stage 2, we do not use weighted Euclidean distance as our metric but instead the "MAP metric" – i.e.,

$$\mathbf{c}^* = \arg\min \left\{ \sum_{j=0}^{44} (y_j - a_j x_j)^2 - N_0 \ln(P(\mathbf{c})) : \mathbf{c} = [c_0, c_1, \ldots, c_{44}] \text{ is on list.} \right\}. \quad (4)$$

  Note that this decoding algorithm requires knowledge of the prior distribution of the codewords – which can be easily computed from Model A's transition probabilities extracted from the training sequence.

- **SDD 4:** This is a sequence-based near-MAP algorithm that seeks to exploit *both* temporal correlation *and* correlation within a frame. Given that a sequence of codewords is sent over the channel, SDD 4 chooses as its estimate the most probable codeword at each step. More precisely, suppose we observe j (corrupted) frames of LSP's; call them $\underline{\mathbf{Y}}=[\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_j]$, and assume we observe $\underline{\mathbf{Y}} = \underline{\mathbf{y}} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j]$. Then the MAP estimate of the j^{th} transmitted codeword given that we also have perfect channel state information (i.e., we know $\underline{\mathbf{A}} = \underline{\mathbf{a}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_j]$), is

$$\begin{aligned} \hat{\mathbf{X}}_j &= \arg\max_{\mathbf{x} \in \mathcal{C}} P(\mathbf{X}_j = \mathbf{x} | \mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, \ldots, \mathbf{Y}_j = \mathbf{y}_j, \mathbf{A}_1 = \mathbf{a}_1, \ldots, \mathbf{A}_j = \mathbf{a}_j) \\ &= \arg\max_{\mathbf{x} \in \mathcal{C}} f_{\underline{\mathbf{Y}}|\mathbf{X}_j, \underline{\mathbf{A}}}(\underline{\mathbf{y}} | \mathbf{x}, \underline{\mathbf{a}}) P(\mathbf{X}_j = \mathbf{x}). \end{aligned}$$

Here $\mathbf{x}$ represent the BPSK-modulated version of a codeword. If we define the objective function to be maximized by

$$g^{(j)}(\mathbf{x}) = f_{\underline{\mathbf{Y}}|\mathbf{X}_j,\underline{\mathbf{A}}}(\underline{\mathbf{y}}|\mathbf{x},\underline{\mathbf{a}})P(\mathbf{X}_j = \mathbf{x}),$$

then by conditioning on the value of $\mathbf{X}_{j-1}$ it can be shown that

$$g^{(j)}(\mathbf{x}) = \sum_{\mathbf{x}' \in \mathcal{C}} f_{\mathbf{Y}_j|\mathbf{X}_j,\mathbf{A}_j}(\mathbf{y}_j|\mathbf{x},\mathbf{a})P(\mathbf{X}_j = \mathbf{x}|\mathbf{X}_{j-1} = \mathbf{x}')g^{(j-1)}(\mathbf{x}'),$$

which implies the maximum objective function can be computed recursively. Furthermore, define $f^*_{\mathbf{Y}_j|\mathbf{X}_j,\mathbf{A}_j}(\mathbf{y}_j|\mathbf{x},\mathbf{a})$ to be an approximation to $f_{\mathbf{Y}_j|\mathbf{X}_j,\mathbf{A}_j}(\mathbf{y}_j|\mathbf{x},\mathbf{a})$:

$$f^*_{\mathbf{Y}_j|\mathbf{X}_j,\mathbf{A}_j}(\mathbf{y}_j|\mathbf{x},\mathbf{a}) = \begin{cases} f_{\mathbf{Y}_j|\mathbf{X}_j,\mathbf{A}_j}(\mathbf{y}_j|\mathbf{x},\mathbf{a}), & \text{if } \mathbf{x} \in A(\mathbf{y}_j); \\ 0, & \text{otherwise;} \end{cases} \tag{5}$$

where $A(\mathbf{y}_j)$ is the list of N modulated codeword candidates obtained in Stage 1 of the SDD 2 algorithm, given the j^{th} received vector $\mathbf{y}_j$. We then get the simplified objective function:

$$g^{(j)}(\mathbf{x}) = \begin{cases} \sum_{\mathbf{x}' \in A(\mathbf{y}_{j-1})} f^*_{\mathbf{Y}_j|\mathbf{X}_j,\mathbf{A}_j}(\mathbf{y}_j|\mathbf{x},\mathbf{a})P(\mathbf{X}_j = \mathbf{x}|\mathbf{X}_{j-1} = \mathbf{x}')g^{(j-1)}(\mathbf{x}'), & \text{if } \mathbf{x} \in A(\mathbf{y}_j); \\ 0, & \text{otherwise.} \end{cases}$$

This, coupled with the fact that the conditional probabilities $P(\mathbf{X}_j = \mathbf{x}|\mathbf{X}_{j-1} = \mathbf{x}')$ can be computed from the Markov parameters obtained from the Model B training sequence, gives us an iterative decoder that exploits *both* the inter-frame *and* the intra-frame correlation present in the LSP's.

In each algorithm, the main *computational* complexity resides in Stage 2; the amount of computation is proportional to N, the number of candidates. The *storage requirement* (assuming an (n, k) code over $\mathcal{F}_q$) is $q^{n-k}L$ and is determined by the size of the syndrome table used in Stage 1.

VI. Simulation Results

Simulation was used to determine the performance of the proposed decoding algorithms. A block diagram of the proposed system is in Figure 1. The three high-order bits of each of

the ten quantized LSP's were channel encoded using one of the two codes described in Section IV. The outputs of the channel encoders were then BPSK-modulated over either the AWGN channel or the fully interleaved Rayleigh channel. After appropriate demodulation, the signals were decoded with the proposed channel decoders and the decoded LSP's were fed into the CELP decoder for speech reconstruction. Note that the system using the convolutional code requires one frame of decoding delay while the Reed-Solomon code requires none.

A large training sequence consisting of 83,826 frames (about 42 minutes of speech) from the TIMIT speech database [13] was used to estimate the prior LSP distributions needed for the MAP 1, MAP 2, SDD 3, and SDD 4 decoders. The testing sequence consisted of 4753 frames (about 2.2 minutes of speech) – 48 sentences, half uttered by female speakers and half by male speakers from different dialect regions. No speaker appeared in both the training and testing sequences. Thus the approach used in this simulation was to use a single "universal" model – constructed from a very large training sequence – to decode *all* the speech samples. An alternative approach – not considered in this paper – would have been to repeatedly re-train the channel decoder for different speakers.

In evaluating the performance of the various decoders we use two criteria. The first is the average spectral distortion (SD), the most commonly used distortion measure for the LSP's [10]. More specifically,

$$SD = \frac{1}{T} \sum_{j=1}^{T} \left[\int_{-\pi}^{\pi} \left(10 \log_{10} S_j(w) - 10 \log_{10} \hat{S}_j(w) \right)^2 \frac{dw}{2\pi} \right]^{\frac{1}{2}} \quad \text{dB},$$

where $S_j(w)$ and $\hat{S}_j(w)$ are the original and reconstructed spectra associated with frame j, and T is the total number of frames. Roughly speaking, an average spectral distortion of 1 dB or less is equivalent to perceptually transparent encoding of the LSP coefficients [10]. In addition to average spectral distortion, the percentage of outliers – i.e., the fraction of frames with distortion greater than 4 dB – were also compiled during the simulation. It should be noted that the spectral distortion introduced by CELP's scalar quantizer alone (when the channel is noiseless) is around 1.50 dB with 0.08 % of outliers $>$ 4 dB.

The second measure of the decoders' performance is symbol error rate – i.e., the fraction of LSP's the decoder decoded incorrectly.

Tables 2-10 and Figures 2-3 describe the simulation results.

Observations Regarding Results:

- Table 2 describes the performance of the convolutional code over the AWGN channel. We see that the two MAP decoders can – at an SNR as low as $E_s/N_0 = 1$ dB ($E_b/N_0 = 2.25$ dB[3]) – recover the LSP's almost as if the channel was noiseless. (An average spectral distortion of 1.61 dB compared with the best-possible 1.5.) Moreover, at these low channel SNR values, the symbol error rates for MAP 1 and MAP 2 are less than 1%, whereas the symbol error rate for the ML decoder is 10-30%.

- Figure 2 displays the spectral distortion results from Table 2 as a function of E_b/N_0. It's clear that the MAP decoders provide exceptional performance for very noisy channels.

- Table 3 lists the coding gains provided by the convolutional codes for particular values of the average spectral distortion and symbol error rate. We note, for instance, that to decode the LSP's with an average spectral distortion of 2.0 dB, the ML decoder requires a channel that is 2.60 dB "cleaner" than is required by the MAP 2 decoder.

- Table 4 describes the performance of the Reed-Solomon codes over the AWGN channel. For each quantization of the matched filter outputs we generate $L = 64$ codeword estimates; these matched filter outputs are quantized by repeatedly toggling the next-least-likely bits to generate more vectors. This continues until there are $N = 64$ distinct candidates on the list or until we have quantized the matched filter outputs $P_{\max} = 64$ ways (i.e., we've toggled the $b = 6$ least likely bits through their possible values). The rationale for the values $N = L = 64$ is shown in Table 5. For a fixed value of $N = PL$,

[3]Recall that E_b is the average energy *per information bit*. So in this case we have $E_b/N_0 = (1/R)(E_s/N_0)$, where R is the rate of the code – i.e., $R = 2/3$ for the Reed-Solomon code and $R = 3/4$ for the convolutional code.

the values of P and L are varied; it is shown in Table 5 that the values $N = L = 64$ and $P = 1$ provide a performance that is slightly superior to any other choice.

- Figure 3 displays the results from Table 4 as a function of E_b/N_0. We observe that SDD 4 provides exceptional performance at very low SNR.

- Table 6 lists the coding gains provided by the Reed-Solomon codes. We see, for instance, that to decode the LSP's with an average spectral distortion of 2.0 dB, the (near-ML) SDD 1 decoder requires a SNR that is 1.77 dB greater than that required by the (near-MAP) SDD 4 decoder.

- Tables 7-10 provide analogous results for the interleaved Rayleigh fading channel. We see in Tables 7 and 8 that, for the convolutional code, the MAP decoders provide 4.35 - 5.27 dB of coding gain on such channels when compared to ML decoding. Similarly, Tables 9 and 10 show that using SDD 4 to decode the Reed-Solomon codes provides 2.5 - 3.56 dB of coding gain compared with SDD 1.

VII. Listening Tests

Listening tests were undertaken to obtain subjective evaluations of the decoding schemes proposed for the Reed-Solomon encoded system. We only considered the case where the channel is AWGN. We made pairwise comparisons of three decoders: Hard-decision ML decoding, SDD 2, and SDD 4. Before starting, each listener was asked to listen to four sample sentences to "anchor" their perspective. The tests were conducted by playing the same output sentence resulting from the same test conditions using two different decoding schemes. The listeners were asked to indicate which of the two outputs sounded better; they were not told which output corresponded to which decoder. If they failed to discern a noticeable difference between the two outputs, they were given the option to choose "neither". Each listener made twelve comparisons – six pairwise comparisons at a $E_s/N_0 = -3$ dB

($E_b/N_0 = -1.24$ dB) and six pairwise comparisons at a $E_s/N_0 = 0$ dB ($E_b/N_0 = 1.76$ dB). The experiments were performed using four different sentences (two sentences at -3 dB and two sentences at 0 dB) and 50 listeners (36 males and 14 females), resulting in a total number of 100 trials for each test and signal-to-noise ratio (SNR).

The results of the listening tests are shown in Table 11. They indicate:

- SDD 4 was overwhelmingly voted the better decoder over HDML at both SNR's.
- At $E_s/N_0 = -3$ dB, SDD 4 was clearly chosen over SDD 2; this was not the case at 0 dB SNR. This is because, when the channel is very noisy, the advantage SDD 4 gains by exploiting more residual redundancy makes a significant difference; at the higher SNR both decoders yield relatively good results, and the difference is much less noticeable.
- Finally, at $E_s/N_0 = -3$ dB, both SDD 2 and HDML sounded poor, resulting in an "undecided" vote of 62%. However, at 0 dB, nearly all the votes (84 %) were cast in favor of SDD 2 with HDML getting no votes at all.

VIII. Summary and Conclusions

We investigated the problem of reliably transmitting CELP-encoded speech over very noisy channels. We started by characterizing the intra-frame and inter-frame LSP residual redundancies that exist at the output of the CELP coders. We proposed two models for LSP generation. These models suggested that, for every frame of speech, at least one-third of the LSP bits are redundant. We next encoded the LSP's using both convolutional and Reed-Solomon codes for transmission over AWGN and fully interleaved Rayleigh fading channels used in conjunction with BPSK modulation. Soft-decision decoders that exploit the LSP residual correlation in combating the channel noise were introduced. For the case of convolutional encoding, three optimal decoders were implemented: one ML decoder and two MAP decoders. For the case of Reed-Solomon encoding, four decoders were implemented:

one near-ML decoder, one decoder that exploited the ordering property, and two near-MAP decoders. Simulation and listening tests results showed that the decoders offered very good performance; coding gains as high as 5 dB were achieved over soft-decision decoders that *do not* exploit the residual LSP redundancy.

References

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

[2] N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An Approach to Combined Source-Channel Coding," *IEEE Transactions on Information Theory,* vol. 33, pp. 827-838, November 1987.

[3] E. Ayanoglu and R. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders," *IEEE Transactions on Information Theory,* vol. 33, pp. 885-865, November 1987.

[4] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," *IEEE Transactions on Information Theory,* vol. 36, pp. 799-809, July 1990.

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

[6] National Communications System (NCS), "Details to Assist in Implementation of Federal Standard 1016 CELP," *Technical Information Bulletin 92-1,* Office of the Manager, NCS, Arlington, VA 22204-2198.

[7] F. Alajaji, N. Phamdo, N. Farvardin and T. Fuja, "Detection of Binary Markov Sources Over Channels with Additive Markov Noise," submitted to *IEEE Transactions on Information Theory,* April 1994.

[8] D. Chase, "A Class of Algorithms for Decoding Block Codes with Channel Measurement Information," *IEEE Transactions on Information Theory*, Vol. 18, No. 1, January 1973.

[9] N. Farvardin and R. Laroia, "Efficient Coding of Speech LSP Parameters Using the Discrete Cosine Transformation," *Proceedings ICASSP-89*, pp. 168-171, 1989.

[10] R. Laroia, N. Phamdo and N. Farvardin, "Robust and Efficient Quantization of Speech LSP Parameters Using Structured Vector Quantization," *Proceedings ICASSP-91*, pp. 661-664, 1991.

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

[12] K. T. Malone and T. R. Fischer, "Enumeration and Trellis-Searched Coding Schemes for Speech LSP Parameters," *IEEE Transactions on Speech and Audio Processing*, Vol. 1, No. 3, July 1993.

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

LSP Redundancy	Model A			Model B		
	ρ_D	ρ_M	ρ_T	ρ_D	ρ_M	ρ_T
LSP 1	0.6816	0.0000	0.6816	0.6816	0.2765	0.9581
LSP 2	0.4804	0.4415	0.9219	0.4804	0.8469	1.3273
LSP 3	0.7566	0.4425	1.1991	0.7566	0.7624	1.5190
LSP 4	0.7093	0.4303	1.1396	0.7093	0.7529	1.4622
LSP 5	0.3495	0.7184	1.0679	0.3495	0.8986	1.2481
LSP 6	0.3585	0.6287	0.9872	0.3585	0.9367	1.2952
LSP 7	0.6764	0.7575	1.4339	0.6764	0.8144	1.4908
LSP 8	0.4511	0.3521	0.8032	0.4511	0.7990	1.2501
LSP 9	0.2953	0.3840	0.6793	0.2953	0.6224	0.9177
LSP 10	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 1: Redundancy (in bits/frame) results for Models A and B using 83826 frames of the TIMIT speech database.

E_s/N_0	SD (dB)			P_s (%)		
	ML	MAP 1	MAP 2	ML	MAP 1	MAP 2
-3 dB	9.47 (97.30%)	4.58 (42.06%)	4.00 (33.07%)	82.51 %	25.21 %	21.02 %
-2 dB	9.29 (93.52%)	3.36 (26.18%)	2.77 (16.76%)	76.60 %	14.34 %	10.00 %
-1 dB	7.88 (80.84%)	2.31 (10.95%)	2.03 (6.37%)	62.07 %	5.66 %	3.49 %
0 dB	5.35 (51.90%)	1.80 (3.22%)	1.70 (1.71%)	35.63 %	1.63 %	0.89 %
1 dB	2.83 (19.28%)	1.61 (0.80%)	1.61 (0.67%)	10.85 %	0.35 %	0.28 %

Table 2: Average spectral distortion and symbol error rate for convolutional codes over AWGN channel. (Values in parentheses are percentages of outliers > 4 dB.)

Coding Gains	SD (dB)				P_s (%)			
	2.0	2.5	3.0	3.5	1 %	5 %	10 %	15 %
MAP 1 vs ML	+ 2.13	+ 2.45	+ 2.62	+ 2.82	+ 1.94	+ 2.33	+ 2.64	+ 2.87
MAP 2 vs MAP 1	+ 0.47	+ 0.49	+ 0.51	+ 0.51	+ 0.41	+ 0.44	+ 0.43	+ 0.44
Total Gain: MAP 2 vs ML	+ 2.60	+ 2.94	+ 3.13	+ 3.33	+ 2.35	+ 2.77	+ 3.07	+ 3.31

Table 3: Coding gains for convolutional codes over AWGN channels for the same average spectral distortion and for the same symbol error rate.

E_s/N_0	Metric	Uncoded	HDML	SDD 1	SDD 2	SDD 3	SDD 4
−3 dB	SD (dB)	6.94 (90.66%)	7.20 (88.60%)	6.32 (76.46%)	4.96 (55.37%)	3.59 (34.40%)	3.18 (27.20%)
	P_s (%)	40.68 %	44.82 %	36.75 %	26.57 %	17.41 %	14.75 %
−2 dB	SD (dB)	6.38 (84.96%)	6.43 (80.26%)	5.17 (60.84%)	3.83 (37.29%)	2.82 (21.42%)	2.54 (16.20%)
	P_s (%)	34.58 %	37.09 %	26.43 %	17.40 %	10.45 %	8.67 %
−1 dB	SD (dB)	5.70 (76.62%)	5.43 (67.26%)	3.90 (40.93%)	2.80 (20.69%)	2.27 (11.76%)	2.08 (7.99%)
	P_s (%)	28.18 %	28.20 %	16.16 %	9.18 %	5.56 %	4.32 %
0 dB	SD (dB)	4.97 (65.37%)	4.34 (50.04%)	2.77 (21.98%)	2.13 (9.43%)	1.89 (5.18%)	1.80 (3.41%)
	P_s (%)	21.85 %	19.33 %	8.04 %	4.08 %	2.49 %	1.91 %
1	SD (dB)	4.19 (51.09%)	3.26 (31.45%)	2.02 (8.42%)	1.76 (3.49%)	1.68 (1.98%)	1.64 (1.30%)
	P_s (%)	15.88 %	11.34 %	2.99 %	1.49 %	0.90 %	0.73 %
2	SD (dB)	3.46 (36.89%)	2.41 (16.31%)	1.66 (2.25%)	1.60 (1.07%)	1.59 (0.84%)	1.58 (0.69%)
	P_s (%)	10.77 %	5.59 %	0.75 %	0.43 %	0.33 %	0.30 %

Table 4: Average spectral distortion and symbol error rate for Reed-Solomon codes over AWGN channel. (Values in parentheses are percentages of outliers > 4 dB.) $N = 64$, $L = 64$ and $P_{max} = 64$.

P	L	SDD 1		SDD 2		SDD 3	
		E_s/N_0: -2 dB	E_s/N_0: 0 dB	E_s/N_0: -2 dB	E_s/N_0: 0 dB	E_s/N_0: -2 dB	E_s/N_0: 0 dB
1	64	5.17 (60.84 %)	2.77 (21.98 %)	4.14 (41.98 %)	2.25 (11.36 %)	3.44 (31.40 %)	2.05 (7.83 %)
2	32	5.18 (60.80 %)	2.83 (22.73 %)	4.17 (42.66 %)	2.32 (12.62 %)	3.54 (32.63 %)	2.13 (9.11 %)
4	16	5.20 (77.67 %)	2.82 (22.65 %)	4.23 (43.26 %)	2.32 (12.73 %)	3.69 (34.90 %)	2.21 (10.14 %)
8	8	5.19 (61.16 %)	2.83 (22.69 %)	4.30 (44.40 %)	2.41 (14.31 %)	3.61 (34.15 %)	2.16 (9.79 %)
16	4	5.26 (62.18 %)	2.93 (24.59 %)	4.44 (46.56 %)	2.45 (14.96 %)	3.90 (39.03 %)	2.36 (13.15 %)
32	2	5.30 (62.70 %)	2.98 (25.51 %)	4.53 (48.65 %)	2.55 (16.71 %)	4.03 (41.47 %)	2.47 (14.92 %)
64	1	5.40 (63.85 %)	3.09 (27.69 %)	4.59 (49.19 %)	2.71 (19.28 %)	4.12 (42.40 %)	2.60 (17.29 %)

Table 5: Average spectral distortion for Reed-Solomon codes over AWGN channels as L and P vary for fixed $N = LP = 64$. (Values in parentheses are percentages of outliers > 4 dB.)

Coding Gains	SD (dB)				P_s (%)			
	2.0	2.5	3.0	3.5	1 %	5 %	10 %	15 %
HDML vs Uncoded	+ 0.10	- 0.13	- 0.39	- 0.58	+ 0.18	- 0.38	- 0.80	- 1.12
SDD 1 vs HDML	+ 1.69	+ 1.53	+ 1.51	+ 1.50	+ 1.89	+ 1.57	+ 1.47	+ 1.40
SDD 2 vs SDD 1	+ 0.71	+ 0.91	+ 0.99	+ 1.03	+ 0.43	+ 0.78	+ 0.86	+ 0.85
SDD 3 vs SDD 2	+ 0.64	+ 0.87	+ 1.04	+ 1.20	+ 0.52	+ 0.64	+ 0.81	+ 0.94
SDD 4 vs SDD 3	+ 0.42	+ 0.49	+ 0.49	+ 0.62	+ 0.17	+ 0.34	+ 0.31	+ 0.41
Total Gain: SDD 4 vs SDD 1	+ 1.77	+ 2.27	+ 2.52	+ 2.85	+ 1.12	+ 1.76	+ 1.98	+ 2.20
Total Gain: SDD 4 vs HDML	+ 3.46	+ 3.80	+ 4.03	+ 4.35	+ 3.01	+ 3.33	+ 3.45	+ 3.60
Total Gain: SDD 4 vs Uncoded	+ 3.56	+ 3.67	+ 3.64	+ 3.77	+ 3.19	+ 2.95	+ 2.65	+ 2.48

Table 6: Coding gains for Reed-Solomon codes over AWGN channels for the same average spectral distortion and for the same symbol error rate. $N = 64$, $L = 64$ and $P_{max} = 64$.

E_s/N_0	SD (dB)			P_s (%)		
	ML	MAP 1	MAP 2	ML	MAP 1	MAP 2
−2 dB	10.21 (98.84%)	5.26 (51.08%)	4.59 (41.42%)	84.90 %	31.68 %	26.33 %
−1 dB	10.02 (97.71%)	4.32 (38.34%)	3.57 (28.32%)	82.79 %	22.76 %	17.05 %
0 dB	9.57 (95.24%)	3.38 (25.84%)	2.75 (17.16%)	79.06 %	14.28 %	9.49 %
1 dB	8.84 (88.84%)	2.59 (14.34%)	2.18 (8.28%)	71.24 %	7.52 %	4.55 %
2 dB	7.55 (76.72%)	2.09 (7.37%)	1.91 (4.26%)	58.20 %	3.59 %	2.24 %
3 dB	5.79 (57.28%)	1.81 (3.17%)	1.72 (1.91%)	40.32 %	1.46 %	0.87 %
4 dB	4.05 (35.36%)	1.69 (1.60%)	1.65 (0.89%)	22.46 %	0.71 %	0.41 %
5 dB	2.73 (17.28%)	1.62 (0.67%)	1.62 (0.55%)	9.87 %	0.28 %	0.28 %

Table 7: Average spectral distortion and symbol error rate of convolutional codes over Rayleigh channel. (Values in parentheses are percentages of outliers > 4 dB.)

Coding Gains	SD (dB)				P_s (%)			
	2.0	2.5	3.0	3.5	1 %	5 %	10 %	15 %
MAP 1 vs ML	+ 3.70	+ 4.09	+ 4.31	+ 4.48	+ 3.69	+ 4.13	+ 4.40	+ 4.59
MAP 2 vs MAP 1	+ 0.65	+ 0.77	+ 0.76	+ 0.79	+ 0.72	+ 0.70	+ 0.66	+ 0.66
Total Gain: MAP 2 vs ML	+ 4.35	+ 4.86	+ 5.07	+ 5.27	+ 4.41	+ 4.83	+ 5.06	+ 5.25

Table 8: Coding gains for convolutional codes over Rayleigh channels for the same average spectral distortion and for the same symbol error rate.

E_s/N_0	Metric	Uncoded	HDML	SDD 1	SDD 2	SDD 3	SDD 4
−2 dB	SD (dB)	7.38 (93.44%)	7.72 (93.46%)	6.97 (84.67%)	5.64 (64.86%)	4.17 (42.35%)	3.79 (36.99%)
	P_s (%)	46.51 %	51.61 %	43.17 %	32.34 %	22.26 %	19.93 %
−1 dB	SD (dB)	6.99 (90.57%)	7.25 (89.72%)	6.24 (75.08%)	4.77 (51.87%)	3.56 (32.44%)	3.19 (26.11%)
	P_s (%)	42.07 %	46.64 %	35.98 %	24.92 %	16.53 %	14.16 %
0 dB	SD (dB)	6.62 (87.42%)	6.78 (84.54%)	5.38 (63.65%)	3.88 (38.35%)	2.97 (22.95%)	2.66 (17.86%)
	P_s (%)	37.75 %	41.30 %	28.19 %	17.74 %	11.25 %	9.34 %
1 dB	SD (dB)	6.20 (83.21%)	6.20 (77.38%)	4.41 (49.08%)	3.12 (25.75%)	2.47 (14.81%)	2.28 (11.42%)
	P_s (%)	33.29 %	35.47 %	20.37 %	11.54 %	7.04 %	5.87 %
2 dB	SD (dB)	5.75 (77.22%)	5.58 (68.92%)	3.53 (34.62%)	2.54 (16.06%)	2.12 (8.94%)	2.00 (6.59%)
	P_s (%)	29.00 %	29.77 %	13.72 %	6.99 %	4.17 %	3.42 %
3 dB	SD (dB)	5.30 (70.51%)	4.89 (58.73%)	2.86 (22.81%)	2.14 (9.30%)	1.90 (5.18%)	1.84 (4.10%)
	P_s (%)	24.98 %	24.05 %	8.59 %	3.99 %	2.40 %	2.03 %
4 dB	SD (dB)	4.85 (63.04%)	4.19 (47.80%)	2.34 (13.79%)	1.89 (5.26%)	1.76 (2.90%)	1.72 (2.21%)
	P_s (%)	21.22 %	18.56 %	5.00 %	2.25 %	1.35 %	1.07 %
5 dB	SD (dB)	4.46 (55.95%)	3.61 (37.83%)	1.99 (7.58%)	1.74 (2.59%)	1.66 (1.41%)	1.65 (1.26%)
	P_s (%)	17.94 %	14.05 %	2.77 %	1.24 %	0.73 %	0.64 %

Table 9: Average spectral distortion and symbol error rate for Reed-Solomon codes over Rayleigh channels. (Values in parentheses are percentages of outliers > 4 dB.) $N = 64$, $L = 64$ and $P_{max} = 64$.

Coding Gains	SD (dB)				P_s (%)			
	2.0	2.5	3.0	3.5	1 %	5 %	10 %	15 %
HDML vs Uncoded	+ 4.12	+ 2.17	+ 1.27	+ 0.60	+ 5.17	+ 1.62	+ 0.31	- 0.59
SDD 1 vs HDML	+ 3.90	+ 3.73	+ 3.33	+ 3.16	+ 5.15	+ 4.08	+ 3.29	+ 2.98
SDD 2 vs SDD 1	+ 1.41	+ 1.59	+ 1.58	+ 1.54	+ 1.20	+ 1.34	+ 1.39	+ 1.37
SDD 3 vs SDD 2	+ 1.01	+ 1.16	+ 1.26	+ 1.40	+ 0.89	+ 0.95	+ 1.04	+ 1.15
SDD 4 vs SDD 3	+ 0.55	+ 0.52	+ 0.59	+ 0.62	+ 0.40	+ 0.35	+ 0.44	+ 0.44
Total Gain: SDD 4 vs SDD 1	+ 2.97	+ 3.27	+ 3.43	+ 3.56	+ 2.49	+ 2.64	+ 2.87	+ 2.96
Total Gain: SDD 4 vs HDML	+ 6.87	+ 7.00	+ 6.76	+ 6.72	+ 7.64	+ 6.72	+ 6.16	+ 5.94
Total Gain: SDD 4 vs Uncoded	+ 10.99	+ 9.17	+ 8.03	+ 7.32	+ 12.81	+ 8.34	+ 6.47	+ 5.35

Table 10: Coding gains for Reed-Solomon codes over Rayleigh channels for the same average spectral distortion and for the same symbol error rate. $N = 64$, $L = 64$ and $P_{max} = 64$.

E_s/N_0 = −3 dB	SDD 2 vs HDML	SDD 2: 23 %	HDML: 15 %	Neither: 62 %
	SDD 4 vs HDML	SDD 4: 94 %	HDML: 1 %	Neither : 5 %
	SDD 4 vs SDD 2	SDD 4: 73 %	SDD 2: 2 %	Neither: 25 %
E_s/N_0 = 0 dB	SDD 2 vs HDML	SDD 2: 84 %	HDML: 0 %	Neither : 16 %
	SDD 4 vs HDML	SDD 4: 93 %	HDML: 0 %	Neither : 7 %
	SDD 4 vs SDD 2	SDD 4: 34 %	SDD 2: 11 %	Neither: 55 %

Table 11: Reed-Solomon codes over AWGN channels: Listening tests results. $N = 64$, $L = 64$ and $P_{max} = 64$.

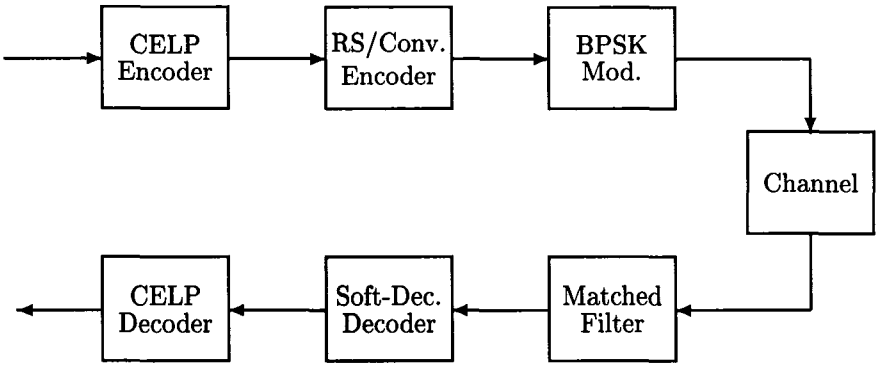


Figure 1: Block diagram of the communication system.

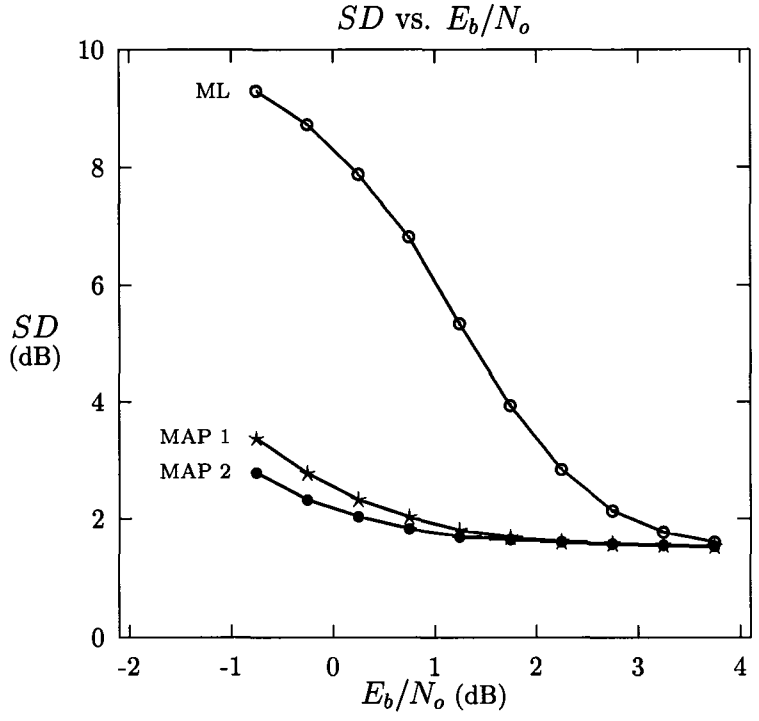


Figure 2: Average spectral distortion for convolutional codes over AWGN channels.

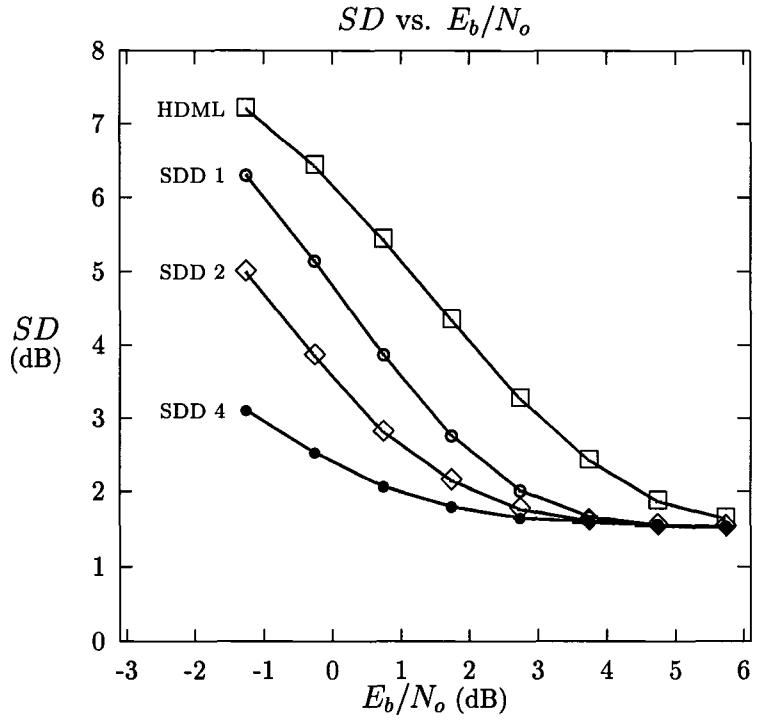


Figure 3: Average spectral distortion for Reed-Solomon codes over AWGN channels. $N = 64$, $L = 64$ and $P_{max} = 64$.