Annual Allerton Conference on Communication, Control and Computing, October 2001.

Sequence MMSE Source Decoding Over Noisy Channels Using the Residual Redundancies

Farshad Lahouti*, Amir K. Khandani

Dept. of E&CE, University of Waterloo
Waterloo, ON, N2L 3G1, Canada
(farshad,khandani)@cst.uwaterloo.ca

Abstract

In this work, we consider the problem of decoding a predictively encoded signal over a noisy channel when there is a residual redundancy (captured by a γ-order Markov model) in the sequence of transmitted data. Our objective is to minimize the mean squared error in the reconstruction of the original signal (input to the predictive source coder). The problem is formulated and solved through Minimum Mean Squared Error (MMSE) decoding of a sequence of samples over a memoryless noisy channel, which was previously recognized to be an open problem by Phamdo and Farvardin in [3]. The related previous works include a sequence MAP decoder [2] and several VQ MMSE decoders which all use a first-order Markov model for the residual redundancy. The former is suboptimal when the performance criterion is the mean squared error and the latter schemes are suboptimal since they decode the data samples received over the channel (the prediction residues) rather than the original signal. As well, using a first-order model, they fail to utilize all the remaining redundancy in the decoding process. The solution is setup by modeling the source and its redundancy with a trellis structure.

1 Introduction

An important result of the Shannon's celebrated paper [1], is that the source and channel coding operations can be separated without any loss of optimality. This has been the basic idea of enormous research endeavors in separate treatment of source and channel coders. However, in practise, there is *residual redundancy* [2] in the output of the source coders which is due to their sub-optimality caused by e.g. a constraint on complexity or delay. As Shannon stated, this redundancy can be used at the receiver to enhance the performance of the system [1].

Recently, researchers have used the residual redundancy for enhanced channel decoding e.g. [5] and [6] or for effective source decoding e.g. [7]-[9]. The problem is formulated in the form of a *Maximum A Posteriori* detection e.g. [3] and [5] or a *Minimum Mean Squared Error* estimation problem e.g. [4]. Several speech error concealment solutions based on MMSE source decoding are presented in [8] and [9]. In [9] a solution to reduce the complexity of the MMSE decoding and its application for error concealment in IS-641 CELP is presented.

*This work is funded in part by the Natural Sciences and Engineering Research Council of Canada.

The recent literature clearly demonstrate the benefit of exploiting the residual redundancies in reconstructing the data received over noisy channels. Although, different applications have been studied, one general problem considered can be viewed as decoding of a source encoded with a memoryless VQ when there is a residual redundancy in the form of a first-order Markov model in the encoder's output sequence. Recently, a general solution for MMSE decoding of a source when the redundancy is captured more efficiently by a γ-order Markov model ($\gamma \geq 1$) was presented in [10].

In this work, we consider the problem of decoding a predictively encoded signal over a noisy channel when there is a residual redundancy (captured by a γ-order Markov model) in the sequence of transmitted data. In fact, in [2] it was shown that there is always a residual redundancy in the output of a DPCM encoder. Our objective is to minimize the mean squared error in the reconstruction of the original signal (input to the predictive source coder). The problem is formulated and solved through Minimum Mean Squared Error (MMSE) decoding of a sequence of samples over a memoryless noisy channel, which was previously recognized to be an open problem by Phamdo and Farvardin in [3]. The solution is setup by modeling the source and its redundancy with a trellis structure. The proposed solution is optimized to minimize the computational complexity. Based on the proposed trellis structure, we also present a Sequence MAP decoder which exploits the redundancies in the form of a γ-order Markov model.

The related previous works include a sequence MAP decoder [2] and several VQ MMSE decoders (e.g. [8]) which all use a first-order Markov model for the residual redundancy. The former is suboptimal when the performance criterion is the mean squared error and the latter schemes are suboptimal since they decode the data samples received over the channel (the prediction residues) rather than the original signal. As well, using a first-order model, they fail to utilize all the remaining redundancy in the decoding process.

The organization of this manuscript is as follows. In section 2, an overview of the system and the channel model used is described. In section 3, the Sequence MMSE decoder is presented. The Sequence MAP decoder is presented in section 4. In section 5, the application of the proposed Sequence MMSE decoding scheme for decoding of the predictively encoded signals is discussed. In section 6, numerical results and comparisons are presented.

2 System Overview

The block diagram of the system is shown in Figure 1. The source encoder $\mathcal{E}$ receives the sample $\mathbf{X}_n$ from an N-dimensional Euclidean space, $\mathcal{R}^N$, and maps it into an index I_n from a finite index set $\mathcal{J}$ of M elements. The source encoder has memory i.e. it may use the previous values of its input and/or output for each mapping. It is composed of two components: the quantizer Q and the index generator $\mathcal{I}$. The quantizer (considering the encoder memory) maps the input $\mathbf{X}_n$, to one of the reconstruction points or *codewords* in $\mathcal{R}^N$ [1]. The bitrate of the quantizer r is given by $\lceil(\log_2 M)\rceil$ bits/symbol (or $\lceil(\log_2 M)\rceil/N$ bits/dim). At the receiver, for each transmitted r-bit index (symbol) $I_n = i_n$, a vector $J_n = j_n$ with r components is received, which depending on the channel model provide

[1]Capital letters (e.g. I) represent random variables, while small letters (e.g. i) is a realization. For simplicity, $P(I = i)$ is represented by $P(I)$. Vectors are shown bold faced (e.g. $\mathbf{X}$). Lower index indicate time instant. Upper index in parenthesis indicate components of a vector or bit positions representing an integer value.

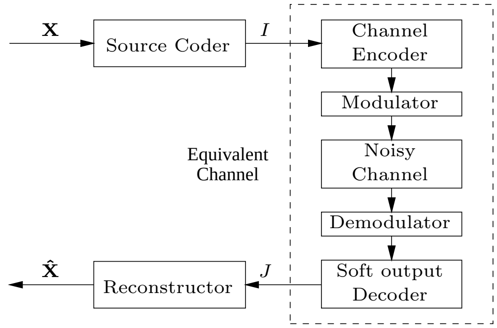


Figure 1: Overview of the system

information about I_n in different ways.The reconstructor (source decoder) uses the information sequence $\underline{J}_{n+\delta} = [J_1, \ldots, J_n, \ldots, J_{n+\delta}]$ and produces an output sample $\hat{\mathbf{X}}_n$. The variable $\delta \geq 0$ denotes the delay allowed in the decoding process.

The noisy channel together with the channel encoder and decoder is replaced by a channel model. We assume that the equivalent channel between I and J is memoryless, i.e.,

$$P(J = j|I = i) = \Pi_{m=1}^{r} P(j^{(m)}|i^{(m)}), \tag{1}$$

where $i^{(m)}, j^{(m)}, m = 1, \ldots, r$ are the bit components of i and j respectively. For a sequence of transmitted symbols, $\underline{I}_n = [I_1, I_2, \ldots, I_n]$ over a memoryless channel, we have,

$$P(\underline{J}_n = \underline{j}_n|\underline{I}_n = \underline{i}_n) = \Pi_{q=1}^{n} P(J_q = j_q|I_q = i_q). \tag{2}$$

Therefore, the channel model includes any abstract memoryless mapping that can be described by a pdf $P(j_n|i_n)$. A basic example of such a channel is the Binary Symmetric Channel. In section 6, we use a BPSK modulation and a channel with AWGN which produces soft outputs.

3 Sequence MMSE Decoding

Consider the case where due to the sub-optimality of the source coder there is a residual redundancy in the output stream of the source coder. This redundancy is in the form of a non-uniform distribution or memory in the sequence of the transmitted symbols. For effective reconstruction of the transmitted information at the receiver, our objective is to exploit this redundancy by designing the reconstructor (source decoder) in Figure 1, such that it produces the Minimum Mean Squared Error estimate of the source sample $\mathbf{X}_n$ given the received sequence $\underline{J}_{n+\delta} = \underline{j}_{n+\delta} = [j_1, j_2, \ldots, j_{n+\delta}]$, where $\delta \geq 0$ is the delay allowed in the decoding process. Based on the fundamental theorem of Estimation Theory, this is given by,

$$\hat{\mathbf{x}}_n = E[\mathbf{X}_n|\underline{J}_{n+\delta} = \underline{j}_{n+\delta}] \tag{3}$$

which minimizes the expected squared error of estimation,

$$E[(\mathbf{X}_n - \hat{\mathbf{X}}_n)'(\mathbf{X}_n - \hat{\mathbf{X}}_n)|\underline{J}_{n+\delta} = \underline{j}_{n+\delta}] \tag{4}$$

The Equation (3) can be expanded as follows,

$$\hat{\mathbf{x}}_n = \sum_{\underline{i}_{n+\delta}} E[\mathbf{X}_n|\underline{I}_{n+\delta} = \underline{i}_{n+\delta}]P(\underline{I}_{n+\delta} = \underline{i}_{n+\delta}|\underline{J}_{n+\delta} = \underline{j}_{n+\delta}). \tag{5}$$

We assume the encoded (quantized) sample $\tilde{\mathbf{X}}_n$ corresponding to the input sample $\mathbf{X}_n$ can be written as a specified function f of the current and the K previously encoded symbols, i.e.,

$$\begin{gathered}\tilde{\mathbf{X}}_n = f(I_{n-K}, \ldots, I_{n-1}, I_n) \\ I_{n-q} \in \mathcal{J}, 0 \leq q \leq K\end{gathered} \tag{6}$$

We refer to an encoder with such a characteristic as an encoder with memory length K. Given the above Equation and assuming that the encoded sequence contain residual redundancy in the form of a Markov model of order γ, $\gamma \geq 1$, the Equation (5) is now simplified to

$$\hat{\mathbf{x}}_n = \sum_{\underline{I}_{n+\delta}^{n-K-\gamma}} E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-K-\gamma}]P(\underline{I}_{n+\delta}^{n-K-\gamma}|\underline{J}_{n+\delta}) \tag{7}$$

where $\underline{I}_{n+\delta}^{n-K-\gamma} = [I_{n-K-\gamma}, \ldots, I_{n+\delta}]$. Equation (7) provides the optimum MMSE reconstruction of the source sample $\mathbf{X}_n$ given the received sequence $\underline{J}_{n+\delta}$ subject to the abovementioned assumptions. This Equation describes the MMSE estimate in terms of the weighted average of the terms $E[\mathbf{X}_n|\underline{I}_{n+\delta}^{n-K-\gamma}]$, which we refer to as the *decoder codewords* or collectively the *decoder codebook*. The weights are the probability of receiving the corresponding sequence of indices given the received sequence $\underline{J}_{n+\delta} = \underline{j}_{n+\delta}$. The decoder codewords provide a finer quantization of the source samples as compared to the encoder codewords which are described as a function of only $\underline{I}_{n-K}^{n}$ (see Equation (6)). This, however, leads to a higher memory requirement at the decoder and assuming that the encoder codebook provides a fine enough quantization of the source, we further simplify the Equation (7) to the following,

$$\hat{\mathbf{x}}_n = \sum_{\underline{I}_n^{n-K} \in \mathcal{J}^{K+1}} E[\mathbf{X}_n|\underline{I}_n^{n-K}]P(\underline{I}_n^{n-K}|\underline{J}_{n+\delta}) \tag{8}$$

which provides the MMSE estimate as a weighted average of the encoder codewords. The weights or the probabilities $P(\underline{I}_n^{n-K}|\underline{J}_{n+\delta} = \underline{j}_{n+\delta})$ are calculated every time instant as developed below. This methodology can also be employed to solve the optimum case (Equation (7)) as discussed in the subsection 3.2.

First, we model the source I_n by a trellis structure. In this structure, the states at time n corresponds to the ordered set

$$S_n = (I_{n-\gamma+1}, I_{n-\gamma+2}, \ldots, I_{n-1}, I_n). \tag{9}$$

Hence, there are M^γ states in each time step (stage), $S_n \in \mathcal{J}^\gamma$. Each branch leaving the state at time step n corresponds to one particular symbol $I_{n+1} = i_{n+1}$. Therefore, there are M branches leaving each state. Each branch is identified by the pair $(S_n = s_n, S_{n+1} = s_{n+1})$ of the two states that the branch connects together. Having defined the trellis structure as such, there will be one a priori probability $P(I_{n+1} = i_{n+1}|S_n = s_n)$ corresponding to each branch which characterizes the γ-order Markov property of the source. The states now form a first-order Markov sequence. Using this property, the memoryless assumption of the channel (see Eq. (1) and (2)), in a similar spirit to the BCJR algorithm [12], the probability of a particular state $S_{n+m} = s_{n+m}$, $0 < m \leq \delta$ given the observed sequence $\underline{J}_{n+\delta}$ is calculated as follows,

$$P(S_{n+m}|\underline{J}_{n+\delta}) = C \,.P(S_{n+m}|\underline{J}_{n+m})\,.\,P(\underline{J}_{n+\delta}^{n+m+1}|S_{n+m}), \quad 0 \leq m \leq \delta \tag{10}$$

where $\underline{J}_{n+\delta}^{n+m+1} = [J_{n+m+1}, J_{n+m+2}, \ldots, J_{n+\delta}]$ and C is a factor which normalizes the sum of probabilities to one. The above equation can be computed using the well known Forward Backward equation [12]. The first term is referred to as the forward equation and is denoted by $\alpha_{n+m}(S_{n+m})$. This can be calculated recursively as follows,

$$\alpha_n(S_n = s_n) = P(J_n = j_n|I_n = i_n)\,. \sum_{s_{n-1}\to s_n} P(I_n = i_n|S_{n-1} = s_{n-1}).\alpha_n(S_{n-1} = s_{n-1}) \tag{11}$$

The second term in Equation (10) is referred to as the backward equation and is denoted by $\beta_{n+m}(S_{n+m})$. This can be calculated recursively as follows,

$$\beta_n(S_n = s_n) = \sum_{i_{n+1}\in\mathcal{J}} P(J_{n+1} = j_{n+1}|I_{n+1} = i_{n+1})\,. P(I_{n+1} = i_{n+1}|S_n = s_n)\,.\,\beta_{n+1}(S_{n+1} = s_{n+1}) \tag{12}$$

Now, the Equation (10) can be rewritten as,

$$P(S_{n+m} = s_{n+m}|\underline{J}_{n+\delta}) = C\,.\alpha_{n+m}(S_{n+m} = s_{n+m})\,.\,\beta_{n+m}(S_{n+m} = s_{n+m}), \quad 0 \leq m \leq \delta \tag{13}$$

where,

$$\beta_{n+\delta}(S_{n+\delta} = s_{n+\delta}) = 1, \quad \forall s_{n+\delta} \in \mathcal{J}^{\gamma} \tag{14}$$

Equations (11)-(14) provide the necessary means to calculate the probabilities of states as described in Equation (10). The weights (probabilities) we need to calculate every time instant to be used in Equation (8) are calculated using the probabilities of states. Depending on the relative value of encoder memory K,to the residual redundancy order γ, this is performed in two ways as described below.

3.1 Calculating the weights for $K < \gamma$

For the scenario with $K < \gamma$, we can calculate the probabilities required in Equation (8), by performing $\gamma - K - 1$ summations over *any of the probabilities* $P(S_{n+m}|\underline{J}_{n+\delta})$ *as long as* S_{n+m} *includes* $\underline{I}_n^{n-K}$ *or equivalently,* $0 \leq m \leq \gamma - K - 1$. However, it is shown that the number of computations required for the forward and backward recursions per time step (denoted by NC_{fwd} and NC_{bwd} respectively) is given by,

$$NC_{fwd} = (2M + 1)\, M^{\gamma} \tag{15}$$

$$NC_{bwd} = 3(\delta - m)\, M^{\gamma+1} \tag{16}$$

where $\delta - m$ is the number of backward recursions required per time step. Therefore, we can select the value of m such that it minimizes the overall computational burden. We solve the following for the optimum value of m,

$$\begin{aligned} &Minimize \quad NC_{bwd} = 3(\delta - m)\,.\,M^{\gamma+1} \\ &subject\ to \quad 0 \leq m \leq \gamma - K - 1; \quad 0 \leq m \leq \delta \end{aligned} \tag{17}$$

case 1. $\delta < \gamma - K$ In the cases where the delay is smaller than the difference of the assumed residual redundancy order and the encoder memory, we are able to choose $m = \delta$ and eliminate the backward term. The probabilities in Equation (8) are calculated using (11) and the following,

$$P(\underline{I}_n^{n-K}|\underline{J}_{n+\delta}) = \ldots \sum_{I_{n+q}} \ldots P(S_{n+\delta}|\underline{J}_{n+\delta}), \quad q = \delta - \gamma + 1, \ldots, \delta, \quad q \neq -K, \ldots, 1, 0. \tag{18}$$

case 2. $\delta \geq \gamma - K$ Alternatively, when the delay is larger than $\gamma - K$, the NC_{bwd} is minimized when $m = \gamma - K - 1$, i.e. $\delta + K - \gamma + 1$ backward recursions is required. The probabilities in Equation (8) are now given by,

$$P(\underline{I}_n^{n-K}|\underline{J}_{n+\delta}) = \sum_{I_{n+1}} \sum_{I_{n+2}} \cdots \sum_{I_{n+\gamma-K-1}} P(S_{n+\gamma-K-1}|\underline{J}_{n+\delta}) \tag{19}$$

and Equations (11) to (14).

3.2 Calculating the weights for $K \geq \gamma$

For the scenario with the residual redundancy order smaller than the encoder memory $K \geq \gamma$, the sequence $\underline{I}_n^{n-K} = [I_{n-K}, \ldots, I_{n-1}, I_n]$ whose a posteriori probability is required, in fact corresponds to a sequence of states within the trellis structure of the source as described before. Consequently, the desired probabilities can be calculated using the probability of the corresponding sequence of states. We have,

$$P(\underline{I}_n^{n-K}|\underline{J}_{n+\delta}) = P(\underline{S}_n^{n-K+\gamma-1}|\underline{J}_{n+\delta}) \tag{20}$$

This can be written in the following forward backward form where we have used the assumption of redundancy order of γ, to replace $P(\underline{J}_{n+\delta}^{n+1}|\underline{S}_n^{n-K+\gamma-1})$ with $P(\underline{J}_{n+\delta}^{n+1}|S_n)$.

$$P(\underline{S}_n^{n-K+\gamma-1}|\underline{J}_{n+\delta}) = C.P(\underline{S}_n^{n-K+\gamma-1}|\underline{J}_n).P(\underline{J}_{n+\delta}^{n+1}|S_n) \tag{21}$$

The value C is a factor which normalizes the sum of probabilities to one. The second term or the backward term is given by the Equations (12) and (14). The forward term is given by,

$$P(\underline{S}_n^{n-K+\gamma-1}|\underline{J}_n) = \Pi_{q=\gamma-K}^{0}[P(J_{n+q}|I_{n+q}).P(I_{n+q}|S_{n+q-1})]P(S_{n-K+\gamma-1}|\underline{J}_{n-K+\gamma-1}) \tag{22}$$

For a comprehensive complexity analysis and alternative implementations of Sequence MMSE Decoder refer to [11]. The Equation (22) can also be used to compute the weights required for the optimum Sequence MMSE decoding given in Equation (7).

4 Sequence MAP Decoding

A Sequence MAP decoder exploiting the residual redundancies in the form of a first-order Markov model was presented in [2]. Later in [3], a similar but optimal Sequence MAP decoder was proposed. Here we present an optimal Sequence MAP decoder when the residual redundancies are captured with a γ-order Markov model.

The Sequence MAP decoder receives the sequence $\underline{J}_n$ and determines the most probable transmitted sequence,

$$\hat{\underline{i}}_n = \arg \max_{\underline{I}_n \in \mathcal{J}^n} P(\underline{I}_n|\underline{J}_n) \tag{23}$$

Using the same trellis structure as described in the previous section and considering the memoryless property of the channel as well as the Markov model for the source redundancy, it is straightforward to see that the Equation (23) is equivalent to,

$$\hat{\underline{i}}_n = \arg \max_{\underline{I}_n \in \mathcal{J}^n} \{ \sum_{q=2}^{n} \log[P(J_q|I_q)P(I_q|S_{q-1})] + P(J_1|I_1)P(S_1) \} \tag{24}$$

The Sequence MAP decoder in Equation (24) can be implemented using the well-known Viterbi algorithm. We use the same trellis structure as defined before and the metric corresponding to branch (S_{q-1}, S_q) is given by $\log[P(J_q|I_q)P(I_q|S_{q-1})]$.

5 Reconstruction of Predictively Encoded Signals

In this section, we consider the MMSE reconstruction of DPCM signals over noisy channels. We focus on the DPCM systems with Auto Regressive prediction. This is due to the popularity of these systems and the fact that the ideas employed in this case can be easily applied to the other cases including Moving Average (linear or nonlinear) predictive encoding systems.

Figure 2, demonstrates the block diagram of a DPCM encoder with Auto Regressive prediction. In this system, the quantized sample $\tilde{\mathbf{X}}_n$ is given by,

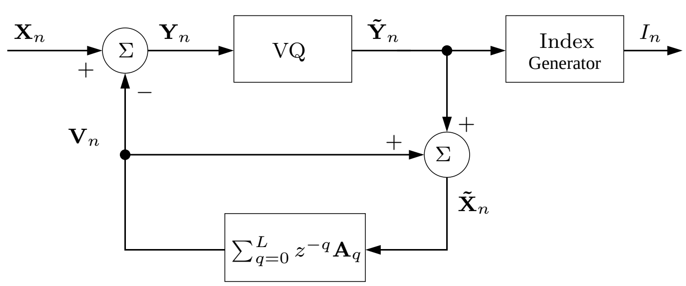


Figure 2: DPCM Encoder

$$\tilde{\mathbf{X}}_n = \tilde{\mathbf{Y}}_n + \sum_{q=1}^{L} \mathbf{A}_q \tilde{\mathbf{X}}_{n-q}. \tag{25}$$

which can be described as a function of the sequence of prediction residues $[\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \ldots, \tilde{\mathbf{Y}}_n]$. Consequently, the Equation (6) holds and a solution based on the proposed Sequence MMSE Decoding exists. However, this implies that the length of the sequence to be decoded grows with time. Since the complexity of the algorithms grow exponentially with the sequence length, this would lead to impractical schemes. A manageable solution is created by defining an *effective memory* length, i.e., assuming that the sample $\tilde{\mathbf{x}}_n$ depends *effectively* on Y_n and only on K previous prediction residue values, $[\tilde{\mathbf{Y}}_{n-K}, \ldots, \tilde{\mathbf{Y}}_{n-1}, \tilde{\mathbf{Y}}_n]$. Therefore, we can finalize the reconstructed value of the residues beyond $n-K$ or equivalently their corresponding output $\mathbf{X}_{n-K-1}$. This idea is supported by the fact that in DPCM systems error in one sample is effectively propagated to a limited number of future samples.

For a first order AR predictive coder, Equation (25) can be rewritten as,

$$\tilde{\mathbf{X}}_n = \tilde{\mathbf{Y}}_n + \mathbf{A}_1 \tilde{\mathbf{X}}_{n-1} = \sum_{q=0}^{n-1} \mathbf{A}_1^q \tilde{\mathbf{Y}}_{n-q}. \tag{26}$$

The MMSE estimate (Equation (8)) is now given by,

$$\begin{aligned}\hat{\mathbf{x}}_n &= E[\mathbf{X}_n|\underline{J}_{n+\delta}] = E[\sum_{q=0}^{n-1}\mathbf{A}_1^q\tilde{\mathbf{Y}}_{n-q}|\underline{J}_{n+\delta}] \\ &= E[\sum_{q=0}^{K}\mathbf{A}^q\tilde{\mathbf{Y}}_{n-q}|\underline{J}_{n+\delta}] + \mathbf{A}^{K+1}\sum_{q=0}^{n-K-2}E[\tilde{\mathbf{Y}}_{n-K-1-q}|\underline{J}_{n+\delta}] \end{aligned} \tag{27}$$

Subsequently, assuming an effective memory length of K, we approximate the second term by $\mathbf{A}^{K+1}\hat{\mathbf{x}}_{n-K-1}$ (replacing $\underline{J}_{n+\delta}$ in the this term with $\underline{J}_{n-K-1+\delta}$). Next, we reach a recursive formula for the Sequence MMSE Decoding tailored for decoding of a first-order AR DPCM system.

$$\hat{\mathbf{x}}_n = E[\sum_{q=0}^{K}\mathbf{A}^q\tilde{\mathbf{Y}}_{n-q}|\underline{J}_{n+\delta}] + \mathbf{A}^{K+1}\hat{\mathbf{x}}_{n-K-1} \tag{28}$$

$$= \sum_{\underline{I}_{n+\delta}^{n-K-\gamma}} E[\sum_{q=0}^{K}\mathbf{A}^q\tilde{\mathbf{Y}}_{n-q}|\underline{I}_{n+\delta}^{n-K-\gamma}]P(\underline{I}_{n+\delta}^{n-K-\gamma}|\underline{J}_{n+\delta}) + \mathbf{A}^{K+1}\hat{\mathbf{x}}_{n-K-1} \tag{29}$$

$$= \sum_{q=0}^{K}\sum_{\underline{I}_{n+\delta}^{n-q-\gamma}} \mathbf{A}^q E[\tilde{\mathbf{Y}}_{n-q}|\underline{I}_{n+\delta}^{n-q-\gamma}]P(\underline{I}_{n+\delta}^{n-q-\gamma}|\underline{J}_{n+\delta}) + \mathbf{A}^{K+1}\hat{\mathbf{x}}_{n-K-1} \tag{30}$$

The solutions provided in Equations (29) and (30) are perfectly implementable. However, further approximation of the decoder codebooks could lead to more efficient systems at a certain level of performance loss. One particular scenario of interest, is the case where the decoder codebook is identical to the encoder codebook. This is derived as below, from Equation (29), by replacing $\underline{I}_{n+\delta}^{n-K-\gamma}$ by $\underline{I}_{n}^{n-K}$ and subsequently approximating the decoder codebook $E[\sum_{q=0}^{K}\mathbf{A}^q\tilde{\mathbf{Y}}_{n-q}|\underline{I}_{n}^{n-K}]$ by $\sum_{q=0}^{K}\mathbf{A}^q E[\tilde{\mathbf{Y}}_{n-q}|I_{n-q}]$.

$$\hat{\mathbf{x}}_n = \sum_{\underline{I}_{n}^{n-K}}\{\sum_{q=0}^{K}\mathbf{A}^q E[\tilde{\mathbf{Y}}_{n-q}|I_{n-q}]\}P(\underline{I}_{n}^{n-K}|J_{n+\delta}) + \mathbf{A}^{K+1}\hat{\mathbf{x}}_{n-K-1} \tag{31}$$

For $K = 0$ the problem collapses to that of the MMSE reconstruction of prediction residues,

$$\hat{\mathbf{x}}_n = \sum_{I_n} E[\tilde{\mathbf{Y}}_n|I_n]P(I_n|\underline{J}_{n+\delta}) + \mathbf{A}\hat{\mathbf{x}}_{n-1} \tag{32}$$

In the next section, we present numerical results for Sequence MMSE decoding of a DPCM signal over noisy channels. We will use Equation (31) for decoding and refer to it as the Sequence MMSE decoder.

6 Numerical Results

To analyze the performance of the proposed decoders, we use a synthesized source similar to [2]. The source here is a tenth-order Gauss-Markov source with the coefficients given in Table 1. The coefficients are matched to the LPC coefficients of a 20ms segment of

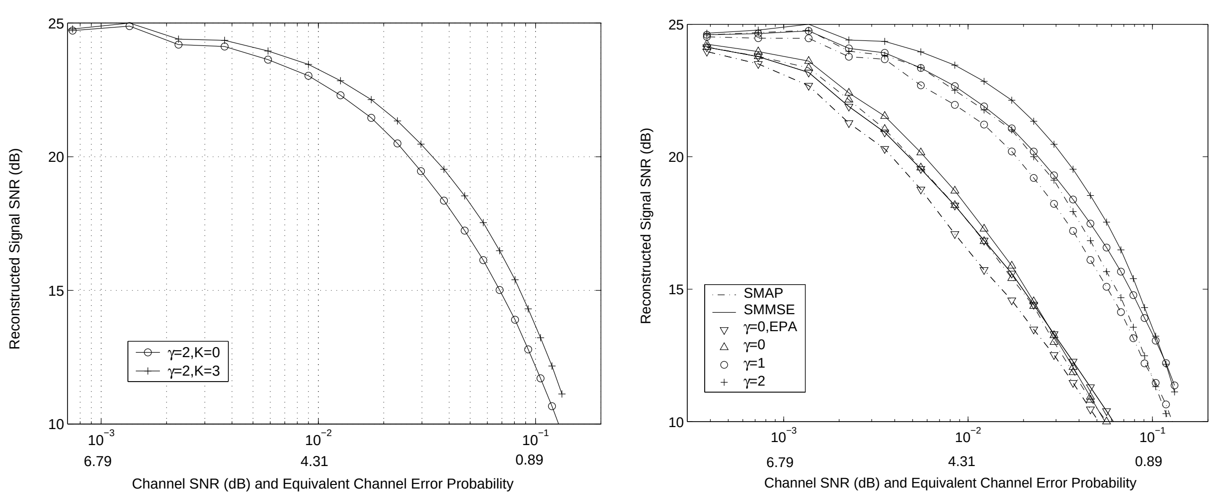
coefficient (1-5)	1.1160	0.5365	-0.1830	-0.5205	-0.0535
coefficient (6-10)	-0.3159	0.3263	-0.0194	0.2841	-0.2006

Table 1: Coefficients of the synthesized source

speech. The source X_n is quantized with a first-order linear auto regressive DPCM coder. The predictor is a noisy channel (Chang and Donaldson) predictor [13]. The quantizer is a Lloyd/Max scalar quantizer [14] with M levels, $M = 8$. The Index Assignment is natural binary. Table 3 depicts the conditional entropy $H(I_n|S_n)$ of the output of the DPCM encoder for different orders of residual redundancy. This provides an indication of the redundancy to be exploited and hence, the gains to be achieved. As given in this Table, the redundancy due to the non-uniform distribution ($\gamma = 0$) is 0.34 bits. The redundancy exploited by means of a first, second and third order Markov model is 1.15, 1.40 and 1.44 bits respectively. Figure 3.a demonstrates the effect of the effective memory length of the decoder K. It is shown that increasing K noticeably enhances the performance. Figure 3.b provides a performance comparison between the proposed Sequence MMSE decoder and the Sequence MAP decoder (section 4) at different orders of residual redundancy. We use a BPSK modulation and an AWGN channel model with soft outputs. It is seen from this figure that the proposed schemes provide effective solutions for source decoding over noisy channels and gain as high as 7dB compared to the Maximum Likelihood decoding.

Redundancy Order γ	0	1	2	3
Conditional Entropy	2.66	1.85	1.60	1.56

Table 3: Conditional Entropy of the source at different values of γ



(a) (b)

Figure 3: Performance of the Sequence MMSE decoder (a) Effect of K or the effective memory length of the decoder ($\gamma = 2, \delta = 0$) (b) Comparison with Sequence MAP decoder ($K = 3, \delta = 0$)(EPA: Equal symbol Probability Assumption)

Also the proposed Sequence MMSE decoder outperforms the sequence MAP decoder by about 2dB.

References

[1] C. E. Shannon, “A mathematical theory of communications,“ *Bell Syst. Tech. J.,* vol. 27, pp. 379-423 and 623-656, 1948.

[2] K. Sayood, J. C. Brokenhagen, “Use of residual redundancy in the design of joint source/channel coders,“ *IEEE Trans. Commun.,* vol.39, No.6, pp. 838-845, 1991.

[3] N. Phamdo and N. Farvardin, “Optimal detection of discrete Markov sources over discrete memoryless channels- Applications to combined-source channel coding,” *IEEE Trans. Inform. Theory,* vol. 40, pp. 186-103, 1994.

[4] D. J. Miller and M. Park, “A sequence-based approximate MMSE decoder for source coding over noisy channels using discrete hidden Markov models,”*IEEE Trans. Commun.,* vol.46, No.2, pp. 222-231, 1998.

[5] F. I. Alajaji, N. Phamdo and T. E. Fuja, “Channel codes that exploit the residual redundancy in CELP-encoded speech,” *IEEE Trans. Speech and Audio Proc.,* vol. 4, No. 5, Sept. 1996.

[6] J. Hagenauer, ”Source-controlled channel decoding,” *IEEE Trans. Commun.,* vol. 43, No. 9, Sept. 1995.

[7] M. Adrat, J. Spittka, S. Heinen and P. Vary, “Error concealment by near optimum MMSE-estimation of source codec parameters,” *IEEE Workshop on Speech Coding,* pp. 84-86, 2000.

[8] T. Fingscheidt and P. Vary, “Softbit speech decoding: A new approach to error concealment,” *IEEE Trans. on Speech and Audio Proc.,* vol. 9, No. 3, Mar. 2001.

[9] F. Lahouti and A. K. Khandani, “Approximating and exploiting the residual redundancies-Applications to efficient reconstruction of speech over noisy channels,” *Proc. IEEE Int. Confs. Acoust., Speech and Sig. Proc.,* vol.2, pp. 721-724, Salt Lake City, UT, May 2001.

[10] F. Lahouti and A. K. Khandani, “An efficient MMSE source decoder for noisy channels,” *Proc. International Symposium on Telecommunications,* pp. 784-787, Tehran, Iran, September 2001.

[11] F. Lahouti, ”Quantization and reconstruction of sources with memory over noisy channels,” (tentative title, under preparation) PhD Thesis, Dept. of E&CE, University of Waterloo, 2001.

[12] L. R. Bahl, J. Cocke, F. Jelinek and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Tran. on Info. Theory,* vol. 20, pp. 284-287, Mar. 1974.

[13] K. Y. Chang and R. W. Donaldson, ”Analysis, optimization, and sensitivity study of differential PCM systems operating on noisy communication channels,” *IEEE Trans. Commun.,* vol. COM-20, pp. 338-350, June 1972.

[14] Y. Linde, A. Buzo, and R.M. Gray, ”An algorithm for vector quantizer design,” *IEEE Trans. Commun.,* vol. COM-28, pp. 84-95, 1980.