APPROXIMATING AND EXPLOITING THE RESIDUAL REDUNDANCIES- APPLICATIONS TO EFFICIENT RECONSTRUCTION OF SPEECH OVER NOISY CHANNELS

Farshad Lahouti, Amir K. Khandani

Dept. of E&CE, University of Waterloo
Waterloo, ON, Canada
farshad@shannon2.uwaterloo.ca, khandani@shannon2.uwaterloo.ca

ABSTRACT

Exploiting the residual redundancy in a source coder output stream during the decoding process has been proven to be a bandwidth efficient way to combat the noisy channel degradations. In this paper, we consider soft reconstruction of LSF parameters in IS-641 CELP coder transmitted over a noisy channel. We propose two schemes. The first scheme attempts to exploit the interframe residual redundancies in the sequence of received parameters. The second approach exploits both interframe and intraframe residual redundancies. Simulation results are provided which demonstrates the efficiency of the algorithms. Another issue addressed here, is a methodology to efficiently approximate and store the residual redundancies or the a priori transition probabilities. For quantizers with high rates calculating these probabilities require a huge number of source samples. As well, storing them require a large amount of memory. These issues can well make the decoder design process an impractical task. The proposed method is based on the classification of the signal domain. The presented schemes provide high quality error concealment solutions for CELP coders.

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, Shannon's work does not put any constraints on the complexity of the systems involved. In practise, there is redundancy in the output of the source coders which is due to their suboptimality 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]. Sayood and Brokenhagen [2], showed that there is always a *residual redundancy* in the output of a DPCM source coder. It was demonstrated that, this is due to certain assumptions need to be made about the data source for system design as well as the structure of the coder itself. They proposed a MAP decoder which showed substantial gains can be achieved when this redundancy is exploited at the receiver. Phamdo and Farvardin [3], suggested MAP decoders as well as an instantaneous MMSE decoder which uses the residual redundancy. Along the same direction, Miller and Park, proposed a sequence-based approximate MMSE decoder [4]. Alajaji et al., [5] also studied exploiting the residual redundancies in FS-1016 LSF coder for improved reconstruction over a noisy channel.

The recent literature clearly demonstrated the benefit of exploiting the residual redundancies in the sequence of noisy data at the receiver. However, beside the redundancy in the sequence of data (in time), the redundancy can exist between different parameters representing a source as well. For example, in speech coding applications using CELP, each frame of speech is represented by a number of parameters. The residual redundancy can both exist in the sequence of certain parameter in time (interframe) and among different parameters representing the speech in one frame (intraframe). In this work, we propose two schemes to exploit the residual redundancies for efficient reconstruction of LSF parameters in IS-641 CELP coder [6] over a noisy channel. The proposed schemes attempt to exploit both intraframe and interframe residual redundancies.

Another issue, for decoders that exploit the residual redundancies is the calculation of a priori transition probabilities. As indicated in [4], specifically for higher quantizer bitrates this will require huge amounts of data. As well, storing them require a large amount of memory.In fact, for practical applications like speech and image coding this can well make the decoder design process an impractical task. In this work, we suggest a method to approximate and store these transition probabilities. Our simulation results prove the accuracy of these approximations.

The organization of this paper is as follows. In section 2, an overview of the system and the channel model used is described. In section 3, the MMSE decoding schemes exploiting the residual redundancies are presented. Also, a methodology to approximate the a priori transition probabilities is proposed. In Section 4, the reconstruction of LSF parameters in IS-641 CELP coder over a noisy channel is studied and numerical results are presented.

2. SYSTEM OVERVIEW

The block diagram of the system is shown in Figure 1. The source encoder $\mathcal{E}$ is a mapping from an N-dimensional Euclidean space, $\mathcal{R}^n$, into a finite index set $\mathcal{J}$ of M elements. It is composed of two components: the quantizer Q and the index generator $\mathcal{I}$. The quantizer maps the input sample $\mathbf{X} \in \mathcal{R}^N$ to one of the reconstruction points or *codewords* from the same space [1]. The index

This work is funded by Natural Science and Engineering Research Council of Canada.

[1]Capital letters (e.g. I) represent random variables, while small letters (e.g. i) is a realization. Vectors are shown bold faced (e.g. $\mathbf{X}$). Lower index indicate time instant and upper index indicate component of a vector or bit positions representing an integer value.

generator is a mapping from the code-book $\mathcal{C}$ to the index set $\mathcal{J}$. The bitrate of the quantizer r is given by $\lceil(\log_2 M)\rceil$ bits/symbol (or $\lceil(\log_2 M)\rceil/N$ bits/dimension). At the receiver, for each trans-

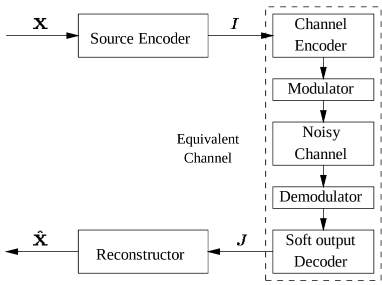

Fig. 1. Overview of the system

mitted r-bit index (symbol) $I = i$, a vector J with r components is received, which depending on the channel model provide information about I in different ways.The reconstructor maps this information to an output sample $\hat{\mathbf{X}}$.

2.1. Channel Model

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,

$$P(\underline{J}_n = \underline{j}_n|\underline{I}_n = \underline{i}_n) = \Pi_{k=1}^{n} P(J_k = j_k|I_k = i_k). \tag{2}$$

The *Binary Symmetric Channel Model* is based on a hard decision on the transmitted bits resulting in a bit error probability of ϵ. In this case, the relationship between the transmitted and the received symbols is given by,

$$P(J = j|I = i) = (\epsilon)^{h_{i,j}}(1-\epsilon)^{(r-h_{i,j})}, \tag{3}$$

where j is the received binary codeword in $\mathcal{J}$ and $h_{i,j}$ is the Hamming distance between indices i and j.

In a system with soft output channel decoding, the channel decoder produces a reliability information vector $J = j$ composed of r Log Likelihood Ratio (LLR) values $[j^1, j^2, \ldots, j^r]$ (see e.g. [7]). In the same direction, the *Soft Output Channel Model* is characterized by using the instantaneous value of $p(j^m|i^m)$ calculated for each bit and equation (1). In this work we use a soft output channel model for an AWGN channel with BPSK modulation.

3. MMSE DECODING EXPLOITING THE RESIDUAL REDUNDANCIES

Based on the fundamental theorem of Estimation Theory, the Minimum Mean Squared Estimate (MMSE) of source sample $\mathbf{X}_n$ given the received sequence $\underline{J}_n = \underline{j}_n = [j_1, j_2, \ldots, j_n]$, is given by,

$$\hat{\mathbf{x}}_n = E[\mathbf{X}_n|\underline{J}_n = \underline{j}_n] \tag{4}$$

which minimizes the expected squared error of estimation,

$$E[\tilde{\mathbf{X}}_n'\tilde{\mathbf{X}}_n] \tag{5}$$

where, $\tilde{\mathbf{X}}_n = \mathbf{X}_n - \hat{\mathbf{X}}_n$. The equation (4), is simplified to

$$\hat{\mathbf{x}}_n = \sum_{i_n \in \mathcal{J}} E[\mathbf{X}_n|I_n = i_n]P(I_n = i_n|\underline{J}_n = \underline{j}_n) \tag{6}$$

which describes the MMSE estimate in terms of the weighted average of LBG codewords. The weights are the probability of receiving the corresponding index given the received sequence $\underline{j}_n$. If we assume that the only element in $\underline{J}_n$ which provides information about I_n is J_n, i.e., there is no residual redundancy, the equation (6) collapses to the basic MMSE reconstruction rule,

$$\hat{\mathbf{x}}_n = \sum_{i_n \in \mathcal{J}} E[\mathbf{X}_n|I_n = i_n]P(I_n = i_n|J_n = j_n) \tag{7}$$

where the probability $P(i_n|j_n)$ contain information about channel condition and the source a priori probabilities $P(i_n)$. We refer to this reconstruction method as MS1 in the following sections. If we assume that the source coder produces equally probable symbols, the term $P(i_n|j_n)$ in equation (7) will be replaced with $P(j_n|i_n)$. We refer to this scenario as MS0.

Assuming the encoded sequence contain residual redundancy [2] in the form of a first order Markov model and a memoryless channel (see Equations (1) and (2)), the probabilities in equation (6) are recursively calculated by,

$$P(I_n = i_n|\underline{J}_n = \underline{j}_n) = C.P(J_n = j_n|I_n = i_n)\,. \sum_{i_{n-1} \in \mathcal{J}} P(I_n = i_n|I_{n-1} = i_{n-1})P(i_{n-1}|\underline{J}_{n-1}) \tag{8}$$

where C is a normalizing constant and $P(i_n|i_{n-1})$ is the $M \times M$ a priori transition probability characterizing the first order Markov residual redundancy. The last term in the above equation is calculated from the same equation in the previous time instant. We refer to this technique as MS2 in the following sections.

The MS2 reconstruction technique, reconstructs a parameter, $\mathbf{X}$ while it exploits the residual redundancy in the sequence of this parameter received over time. However, it neglects any other information that might be available in other received parameters about $\mathbf{X}$. Consider the scenario where the source coder produces two output symbols I and $\tilde{I}$ at a time. Let's assume that due to complexity constraints, there is some level of correlation between the two symbols. At the receiver, this information can be exploited to enhance the reconstruction performance. The MMSE reconstruction (4) is now given by,

$$\hat{\mathbf{x}}_n = E[\mathbf{X}_n|\underline{J}_n = \underline{j}_n, \underline{\tilde{J}}_n = \underline{\tilde{j}}_n] \tag{9}$$

which can be approximated to

$$\hat{\mathbf{x}}_n = \sum_{i_n \in \mathcal{J}} E[\mathbf{X}_n|I_n = i_n]P(I_n = i_n|\underline{J}_n = \underline{j}_n, \underline{\tilde{J}}_n = \underline{\tilde{j}}_n) \tag{10}$$

The probabilities in this equation can be calculated using the forward/backward algorithm [8] (see also [4]). The forward recursions, here is from J_1 to J_n and the backward recursions from J_n

to $\tilde{J}_n$ and from there to $\tilde{J}_1$. To keep the complexity at a manageable level, we approximate these probabilities with the following,

$$P(I_n = i_n|\underline{J}_n = \underline{j}_n, \tilde{J}_n = \tilde{j}_n)) = C.P(j_n|i_n) \, . \sum_{i_{n-1} \in \mathcal{J}} P(I_n = i_n|I_{n-1} = i_{n-1})P(I_{n-1} = i_{n-1}|\underline{J}_{n-1}) \, . \sum_{\tilde{i}_n \in \mathcal{J}} P(\tilde{J}_n = \tilde{j}_n|\tilde{I}_n = \tilde{i}_n)P(\tilde{I}_n = \tilde{i}_n|I_n = i_n) \qquad (11)$$

which can be calculated with only one round of backward recursion. We refer to this technique as MS3 in the following sections.

Employing the derived equations to calculate the instantaneous MMSE estimate require the a priori transition probabilities $P(i_n|i_{n-1})$ and $P(\tilde{i}_n|i_n)$. These are matrices of size $M \times M$ and $M \times \tilde{M}$. For the encoders with high rates, this can be a challenging task since a very large source database is required. In the next section, we propose a method to approximate these transition probabilities.

3.1. Approximating the A Priori Transition Probabilities

Consider the codebook of the encoder $\mathcal{E}$ consisting of M elements. We are interested to find the probability of occurring a certain codeword i_n given that the previous codeword is i_{n-1}. To derive this, we intend to classify the source (and hence the codebook) to $M', M' \le M$, classes in a way that we can make the assumption that the probability of transition from a codeword to another codeword only depends on the class they are located in. Should we have the codebook classified in this manner, we would only need $M' \times M'$ transition probabilities to characterize the residual Markov property of the encoder output sequence. Subsequently, we can derive the $M \times M$ transition probabilities by the following simple derivations. Let's assume that the codebook is classified to M' classes denoted by $\mathcal{J} = \{\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_{M'}\}$. Each class $\mathcal{J}_k$ has m_k members ($\sum_{k=1}^{M'} m_k = M$). We have,

$$\begin{aligned} P(I_n \in \mathcal{J}_k|I_{n-1} \in \mathcal{J}_l) &= \sum_{i_n \in \mathcal{J}_k} P(I_n = i_n|I_{n-1} \in \mathcal{J}_l) \\ &= \sum_{i_n \in \mathcal{J}_k} \sum_{i_{n-1} \in \mathcal{J}_l} P(i_n|I_{n-1} = i_{n-1})P(i_{n-1}|I_{n-1} \in \mathcal{J}_l) \\ &= m_k \, . \, P(I_n = i_n|I_{n-1} = i_{n-1}) \end{aligned} \qquad (12)$$

which shows that the codeword transition probabilities can be approximated as a scaled version of the class transition probabilities. The last result is derived using the assumption mentioned above, i.e.

$$\begin{aligned} &P(I_n = i_n|I_{n-1} = i_{n-1}) = \\ &\quad P(I_n = i_n|I_{n-1} = i'_{n-1}) = P_{kl}, \\ &\quad \forall i_n, i'_n \in \mathcal{J}_k, \forall i_{n-1}, i'_{n-1} \in \mathcal{J}_l \end{aligned} \qquad (13)$$

In order to classify the codebook in a way such that the equation (13) holds, we propose LBG [11] quantization of the source with M' levels and defining the classes as the quantization Voronoi regions. Subsequently, we can classify the codewords of the size M codebook. Therefore, the problem now collapses to that of determining a transition probability matrix of size $M' \times M'$, which can be found with much less data. As well, since the equation (12) provides a simple way to calculate these probabilities from class transition probability, we can store the class transition probabilities instead. This will reduce the memory size needed by a factor of $(\frac{M}{M'})^2$. The transition probabilities $P(\tilde{I}_n|I_n)$ in equation (11) can be approximated in a similar fashion.

4. RECONSTRUCTION OF LSF PARAMETERS

In this section, we study the application of the proposed MMSE decoding schemes for reconstruction of LSF [9] parameters in IS-641 CELP coder [6] over a noisy channel. In this coder, first order Moving Average scalar linear prediction is employed to exploit the redundancies between the adjacent frames. Although, there exists interframe dependencies beyond the adjacent frames, the prediction is limited to one frame to keep the propagation of channel errors at a low level of one frame. Therefore, one can see that there will remain a residual redundancy between the prediction residues which is mainly due to the low prediction order. We refer to these dependencies as the interframe residual redundancies.

The LSF prediction residues in IS-641 is quantized using a 3-split Split-VQ [10] of dimensions $[3, 3, 4]$ with bitrates $[8, 9, 9]$ with an overall rate of 26 bits/frame. Although the encoder attempts to exploit the intraframe dependencies of LSF parameters, due to its suboptimality, there will remain some level of dependency between the quantizer's 3 output symbols. We refer to these dependencies as the intraframe residual redundancies.

4.1. Experimentation Setup

In this work, we use a training database of $175,726$ LSF vectors derived from a 58.57 minute long recorded speech (20ms frame). Another outside test database of $30,000$ LSF vectors is used to test the performance of the system[2]. The spectral distortion measure [10] (measured in frequency range of 60 Hz to 3500 Hz) is employed to measure the objective distortion introduced in the LPC coefficients reconstructed over a noisy channel.

The required a priori transition probabilities are described with one $2^8 \times 2^8$ matrix for the first split and two $2^9 \times 2^9$ probability transition matrices for the other two splits. One can see that finding these values requires a very large speech database which can make the task impractical. As well, the memory required to store these matrices is more than 2 Megabytes. Therefore, approximation of these values is inevitable. Using the method described in section 3.1 with a class size of $M' = 32$ for each of the splits, the problem is reduced to calculating three $2^5 \times 2^5$ transition probability matrices and only 12 kilobytes of memory requirement which is perfectly manageable.

As well, we use the same technique to approximate the transition probabilities between splits. In the following sections, we will use these approximate values wherever the residual redundancy transition probabilities are required and we will study the performance of the system.

4.2. Performance Evaluation

Five decoding schemes are considered here. First a basic Maximum Likelihood (Hard Decision Decoding) algorithm is considered for reference. The methods MS0 and MS1 are the basic MMSE reconstruction algorithms neglecting all the residual redundancies with MS1 exploiting the symbol a priori information $P(i_n)$ (equation (7)). The method MS2 exploit only interframe redundancies for reconstruction (equations (6) and (8)). We employ

[2] These databases were provided by Nortel Networks.

the method MS3 to exploit both intraframe and interframe residual redundancies at the receiver (equations (10) and (11)). In this scheme, to reconstruct the LSF parameters of the first split the information about the received symbol of the second split is used and vice versa. To reconstruct the LSF parameters of the third split the information about the received symbol of the second split is used. This selection is due to the fact that the intraframe correlations of LSF parameters are higher between the neighboring parameters. Tables 1, 2 and 3 depict the performance of the above mentioned schemes, for reconstruction of IS-641 encoded LSF parameters in various channel conditions. It is clear from the tables that employing the a priori information, interframe and intraframe residual redundancies constantly improves the reconstruction quality. The method, MS3 is constantly better than MS2, MS2 is constantly better than MS1 and MS1 is constantly better than MS0 even in good channel conditions. In poor channel conditions, the MMSE decoders, rely more on the source a priori information rather than the information received from the channel. Therefore, the performance advantage of MS1, MS2 and MS3 is higher in these channel conditions. The advantage of exploiting the intraframe residual redundancies are higher compared to that of interframe redundancies and this is due to the fact that source encoder contains first order MA prediction which exploits the interframe correlation and reduces the residual information. The provided results validate the approximate method proposed for calculation of the a priori transition probabilities.

SNR	BER	ML	MS0	MS1	MS2	MS3
1.00	0.0560	4.66	4.26	3.63	3.51	3.16
2.00	0.0370	3.76	3.55	3.04	2.94	2.65
3.00	0.0220	2.87	2.81	2.44	2.36	2.14
4.00	0.0120	2.14	2.13	1.89	1.84	1.69
5.00	0.0059	1.58	1.60	1.47	1.44	1.37
6.00	0.0023	1.25	1.26	1.20	1.19	1.14
7.00	0.0008	1.08	1.09	1.08	1.07	1.06

Table 1. Spectral Distortion [dB] of the test LSF database decoded over noisy channel (SNR 1 to 7) using five decoding schemes.

SNR	BER	ML	MS0	MS1	MS2	MS3
1.00	0.0560	85.51	90.58	82.60	81.27	75.63
2.00	0.0370	71.85	78.86	68.87	67.13	60.63
3.00	0.0220	53.65	60.08	50.95	48.94	42.96
4.00	0.0120	34.98	38.76	32.35	30.85	26.48
5.00	0.0059	19.39	21.12	17.65	16.91	14.48
6.00	0.0023	9.62	10.19	8.38	8.12	7.15
7.00	0.0008	4.44	4.63	4.23	4.13	3.70

Table 2. 2dB Outliers [%] of the test LSF database decoded over noisy channel (SNR 1 to 7) using five decoding schemes.

5. CONCLUSIONS

Different methods were presented for high quality MMSE decoding of signals transmitted over a noisy channel. The residual redundancies in the source coder output stream is exploited at the receiver. These methods were applied to reconstruction of LSF parameters in IS-641 CELP coder transmitted over noisy channel with soft output decoding. An efficient method for approximation and storing the residual redundancies based on classification of the source signal is presented.

SNR	BER	ML	MS0	MS1	MS2	MS3
1.00	0.0560	52.84	46.51	33.81	31.19	23.95
2.00	0.0370	37.95	31.85	22.73	20.82	15.19
3.00	0.0220	23.84	19.26	13.48	12.49	8.86
4.00	0.0120	13.56	10.36	7.11	6.45	4.50
5.00	0.0059	6.47	4.78	3.21	2.85	2.05
6.00	0.0023	2.60	1.82	1.26	1.12	0.77
7.00	0.0008	0.75	0.59	0.39	0.37	0.23

Table 3. 4dB Outliers [%] of the test LSF database decoded over noisy channel (SNR 1 to 7) using five decoding schemes.

6. 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 Transactions on Speech and Audio Processing, vol. 4, No. 5, Sept. 1996.

[6] “TDMA Radio Interface, Enhanced Full-Rate Speech Codec,” TIA/EIA PN-3467, Feb. 1996.

[7] J. Bakus and A. K. Khandani, “Combined source-channel coding using turbo-codes,” *Electronics Letters*, vol. 33, No. 13, Sep. 1997, pp. 1613-1614.

[8] L. E. Baum, T. Petrie, G. Soules and N. Weiss, “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains,” Ann. Math. Stat., vol. 41, pp. 164-171, 1970.

[9] F. Itakura, ”Line spectrum representation of linear predictive coefficients of speech signals,” *Journal Acoustical Society America*, vol.57, p. 535, Apr.1975

[10] K. K. Paliwal and B.S. Atal, ”Efficient vector quantization of LPC parameters at 24 bits/frame,” *IEEE Trans. Speech and Audio Process.*, vol.1, no.1, pp. 3-14, 1993.

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