RECONSTRUCTION OF MULTI-STAGE VECTOR QUANTIZED SOURCES OVER NOISY CHANNELS- APPLICATIONS TO MELP CODEC

Farshad Lahouti, Amir K. Khandani

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

ABSTRACT

Design of source decoders that employ the residual redundancy at the source coder output is an interesting research direction in the joint source channel coding framework. Such decoders are expected to replace the traditionally heuristic error concealment units that are elements of most multimedia communication systems. In this work, we consider the reconstruction of signals encoded with a Multi-Stage Vector Quantizer and transmitted over a noisy channel. The MSVQ maintains a moderate complexity and, due to its successive refinement feature, is a suitable choice for the design of layered (progressive) source codes. An approximate MMSE source decoder for MSVQ is presented and its application to reconstruction of LPC parameters in MELP is analyzed. Numerical results demonstrates the effectiveness of the proposed schemes.

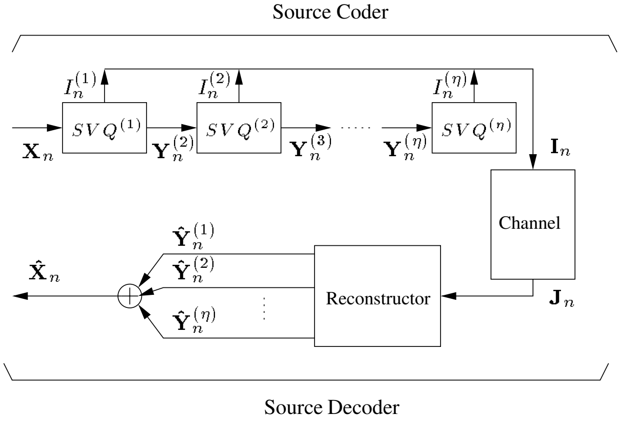

Fig. 1. Overview of the system

1. INTRODUCTION

Recently, methods to exploit the *residual redundancy* [1] in the output of a source coder for improved reconstruction of a signal transmitted over error prone channels has found increasing attention, e.g., [1]-[9]. As a method of joint source channel coding, researchers have used the residual redundancy for enhanced channel decoding, for effective source decoding, or for iterative source and channel decoding. The residual redundancy is often modeled by Markov models.

Source decoders that exploit the residual redundancy provide effective new solutions for concealment of errors in multimedia communications. In this direction, applications to decoding of compressed speech are presented in, e.g., [2], [3]. Applications to robust transmission of digital images over noisy channels are presented in, e.g., [4][5].

In this work, we investigate the redundancy at the output of a Multi-Stage Vector Quantizer and, present an approximate minimum mean squared error technique for reconstruction of MSVQ-encoded sources transmitted over a noisy channel. Numerical results are presented for application of the proposed techniques to reconstruction of LSF parameters in Mixed Excited Linear Prediction (MELP)[10] speech codec.

Due to its structure, Multi-Stage Vector Quantizer benefits from a moderate level of complexity and a reduced codebook size at the cost of a suboptimal performance, when compared to a full-search vector quantizer. One important characteristic of the Multi-Stage Vector Quantizer, also known as the Residual Vector Quantizer, is its (additive) successive refinement feature. This characteristic makes it attractive for designing progressive source codes for effective multimedia communications, in presence of error/loss and delay. For a comprehensive review of the applications of MSVQ in image coding refer to [11]. The MSVQ is part of the MELP speech coding standard [10] and it is applied to video coding in [12].

The rest of this article is organized as follows. In section 2, notations are introduced and the MSVQ structure is briefly described. An MMSE-based algorithm for the reconstruction of MSVQ encoded signals over noisy channels is presented in section 3. Applications to Multi-Stage Vector Quanitzation of LSF parameters in MELP and numerical results are presented in section 4.

2. MULTI-STAGE VECTOR QUANTIZATION

2.1. Preliminaries

In this paper, 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. The sequence of random variables $\{I_{n_1}, \ldots, I_{n_2}\}$ over time is represented by $\underline{I}_{n_2}^{n_1}$, and when $n_1 = 1$ simply by $\underline{I}_{n_2}$.

Figure 1 depicts the block diagram of the system. The source coder is an MSVQ, that is composed of η stages. We refer to each

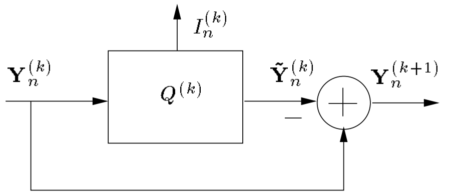

Fig. 2. The kth stage (Stage-VQ) of the MSVQ ($SVQ^{(k)}$)

stage of the MSVQ as an *Stage-VQ (SVQ)*. The input to the system at time n is the signal $\mathbf{X}_n \in \mathcal{R}^N$.

The structure of the SVQ is presented in Figure 2. At time instant n, the input to the k'th SVQ, $1 \leq k \leq \eta$, is the signal $\mathbf{Y}_n^{(k)} \in \mathcal{R}^N$, that following a vector quantization operation, is mapped to two outputs: (i) the index $I_n^{(k)}$ within the finite index set $\mathcal{J}^{(k)}$ of $M^{(k)}$ elements, which corresponds to the codeword $\tilde{\mathbf{Y}}_n^{(k)}$ of the quantizer $Q^{(k)}$, and (ii) the quantization error signal $\mathbf{Y}_n^{(k+1)} \in \mathcal{R}^N$. This signal is the input to the subsequent SVQ. We have

$$\mathbf{Y}_n^{(1)} = \mathbf{X}_n, \tag{1}$$

$$\mathbf{Y}_n^{(k)} = \mathbf{X}_n - \sum_{i=1}^{k-1} \tilde{\mathbf{Y}}_n^{(i)}, \qquad 1 < k \leq \eta. \tag{2}$$

The source coder output is the symbol

$$\mathbf{I}_n = [I_n^{(1)}, I_n^{(2)}, \ldots, I_n^{(\eta)}],$$

composed of the output indices of different Stage-VQs. To search the MSVQ codebook, we use the so-called M-best search as described in [13]. Corresponding to the MSVQ input signal $\mathbf{X}_n$, the MSVQ output quantized signal is given by

$$\tilde{\mathbf{X}}_n = \sum_{k=1}^{\eta} \tilde{\mathbf{Y}}_n^{(k)}. \tag{3}$$

The bitrate of the k'th SVQ, $r^{(k)}$, is given by $\lceil(\log_2 M^{(k)})\rceil$ bits per symbol or $\lceil(\log_2 M^{(k)})\rceil / N$ bits per dimension. The MSVQ bitrate is given by $r = \sum_{k=1}^{\eta} r^{(k)}$.

At the receiver, for each transmitted r-bit symbol $\mathbf{I}_n$, a vector $\mathbf{J}_n$ with r components is received. The source decoder maps this information to an output sample $\hat{\mathbf{X}}_n$.

The noisy channel together with the channel encoder and decoder is replaced by a channel model. We assume that the equivalent channel between $I_n^{(k)}$ and $J_n^{(k)}$ is memoryless, and the probability $P(J_n^{(k)} | I_n^{(k)})$ is available at the source decoder. For simulations in section 4, we assume a BPSK modulation and an AWGN channel, which produces soft outputs.

2.2. Redundancy in MSVQ

For a Multi-Stage Vector Quantizer Residual redundancy could exist in different forms. Since the k'th stage of the MSVQ quantizes a quantization error signal, to analyze the MSVQ one requires to know the statistical behaviour of the quantization error. Lee and Neuhoff presented a high resolution analysis of the error density of vector quantization in [14]. For stationary sources, they showed that, the marginals of the multidimensional error density of an optimal vector quantizer with large dimension are approximately i.i.d. Gaussian. In another line of works on analysis of uniform scalar quantization, it is shown that only under certain strict conditions, the quantization noise is uniform, independently distributed and uncorrelated with the quantizer input [15][16]. As a result, in general, the residual redundancy in a MSVQ could exist in the form of non-uniform symbol probability, and dependency between different stage quantizer outputs. A correlated MSVQ input signal results in residual redundancy in the sequence of output symbols over time.

3. RECONSTRUCTION OF MSVQ-ENCODED SIGNALS

As discussed, a Multi-Stage Vector Quantizer, like many other practical source coders, leave some level of redundancy in its output stream. Our objective is to design a source decoder that exploits this residual redundancy and produces the minimum mean squared error estimate of the source sample $\mathbf{X}_n$, given the received sequence $\underline{\mathbf{J}}_n = [\underline{J}_n^{(1)}, \underline{J}_n^{(2)}, \ldots, \underline{J}_n^{(\eta)}]$. Based on the fundamental theorem of estimation, this is given by $\hat{\mathbf{x}}_n = E[\mathbf{X}_n | \underline{\mathbf{J}}_n]$, and equivalently

$$\hat{\mathbf{x}}_n = \sum_{\underline{\mathbf{I}}_n} E[\mathbf{X}_n | \underline{\mathbf{I}}_n] P(\underline{\mathbf{I}}_n | \underline{\mathbf{J}}_n). \tag{4}$$

In equation (4), the decoder codebook $E[\mathbf{X}_n | \underline{\mathbf{I}}_n]$ provides a finer reconstruction of the source, when compared to the quantized signal at the encoder. This comes at the cost of extra memory requirement. To maintain a reasonable level of complexity, we choose to use the same encoder codebook at the decoder. This is equivalent to assuming $E[\mathbf{X}_n | \underline{\mathbf{I}}_n] \approx \tilde{\mathbf{X}}_n$. Now, using equation (3) and given that $\tilde{\mathbf{Y}}_n^{(k)}$ is specified by $I_n^{(k)}$, the equation (4) is simplified to

$$\hat{\mathbf{x}}_n \approx \sum_{k=1}^{\eta} \sum_{I_n^{(k)} \in \mathcal{J}^{(k)}} \tilde{\mathbf{Y}}_n^{(k)} P(I_n^{(k)} | \underline{\mathbf{J}}_n). \tag{5}$$

This indicates a source decoder that has the same structure as depicted in figure 1, i.e., a reconstructor followed by a summation unit. Note that, in traditional MSVQ decoding the reconstructor is simply a mapping of the received indices to corresponding codewords. In equation (5), the a posteriori probability

$$P(I_n^{(k)} | \underline{\mathbf{J}}_n) = P(I_n^{(k)} | \underline{J}_n^{(1)}, \ldots, \underline{J}_n^{(\eta)})$$

encapsulates the dependencies of symbol $I_n^{(k)}$ with other MSVQ symbol stages at time instant n (intraframe), as well as MSVQ output symbols of previous time instants (interframe). Methods to calculate similar a posteriori probabilities under various assumptions for the redundancy model and using different approximations or formulations are discussed in [3] [6] [7].

To maintain a low level of complexity, we devise the following approximation:

$$\hat{\mathbf{x}}_n \approx \sum_{k=1}^{\eta} \sum_{I_n^{(k)}} \tilde{\mathbf{Y}}_n^{(k)} P(I_n^{(k)} | \underline{J}_n^{(k)}, J_n^{(k')}, J_n^{(k'')}), \tag{6}$$

which exploits the dependency of each symbol, $I_n^{(k)}$, with only two other symbols, $I_n^{(k')}$ and $I_n^{(k'')}$, $k', k'' \neq k \in \{1, \ldots, \eta\}$, at the same time instant n and also with the same symbol at the previous time instant. The parameters k' and k'' are selected for each k, based on the particular application scenario and the available intraframe dependencies. Assuming a first-order Markov model to capture both the interframe and intraframe dependencies and a memoryless channel, the required a posteriori probability in equation (6) is calculated by

$$P(I_n^{(k)}|\underline{J}_n^{(k)}, J_n^{(k')}, J_n^{(k'')}) \approx \\ C.P(I_n^{(k)}|\underline{J}_n^{(k)}).P(J_n^{(k')}|I_n^{(k)}).P(J_n^{(k'')}|I_n^{(k)}), \qquad (7)$$

where,

$$P(I_n^{(k)}|\underline{J}_n^{(k)}) = C'.P(J_n^{(k)}|I_n^{(k)}) . \\ \sum_{I_{n-1}^{(k)} \in \mathcal{J}^{(k)}} P(I_n^{(k)}|I_{n-1}^{(k)}) P(I_{n-1}^{(k)}|\underline{J}_{n-1}^{(k)}) \qquad (8)$$

is computed recursively over time, and,

$$P(J_n^{(k')}|I_n^{(k)}) = \\ \sum_{I_n^{(k')} \in \mathcal{J}^{(k')}} P(J_n^{(k')}|I_n^{(k')}) P(I_n^{(k')}|I_n^{(k)}). \qquad (9)$$

In equations (7) and (8), C and C' are terms that normalize the sum of the probabilities to one. As well, the transition probabilities $P(I_n^{(k')}|I_n^{(k)})$ and $P(I_n^{(k)}|I_{n-1}^{(k)})$, respectively represent the interframe and intraframe dependencies and are stored at the decoder.

4. PERFORMANCE ANALYSIS

In this section, we consider the application of the presented MMSE-based decoder for MSVQ-encoded signals to reconstruction of Line Spectral Parameters in MELP [10]. We use a training database of $175,726$ LSF (20ms frame). This database contains a combination of clean speech and speech with background noise from a number of male and female speakers. Another outside test database of $30,000$ LSF vectors derived from recorded clean speech is used to test the performance of the quantizers[1]. The spectral distortion measure (measured in the frequency range of $f_1 = 60$ Hz to $f_2 = 3500$ Hz) is employed to measure the objective quality of the reconstructed LPC coefficients.

4.1. MSVQ for Quantization of LSF Parameters

In the 2.4 kb/s Mixed Excitation Linear Prediction speech codec [10], that is selected as a U.S. Federal Standard in 1997, a Multi-Stage Vector Quantizer is used for quantization of speech Linear Prediction Coefficients in the Line Spectral Frequency representation. The linear prediction order is 10 and therefore, the number of LSF parameters in each frame and subsequently, the VQ dimension in MSVQ is also 10. The MSVQ in this standard consists of 4 stages, with bit-rates of 7,6,6,6 bits for an overall rate of 25 bpf. An 8-best search is used to find the nearest codewords.

In this work, we use a slightly different structure for Multi-Stage Vector Quantization of LSF parameters. This structure of MSVQ consists of 4 stages of 64 codevectors (6 bits) each, for an overall bitrate of 24 bpf. For code-book search a 2-best search is used. In fact, this configuration is recognized in [13] as "one of the best [MSVQ] configurations in terms of the trade-off between complexity and performance".

Stage No.	1	2	3	4
1	5.77	0.66	0.41	0.35
2	0.66	5.84	0.31	0.24
3	0.41	0.31	5.89	0.26
4	0.35	0.24	0.26	5.92

Table 1. Mutual information of different MSVQ stages within a frame in bits, diagonal elements of the table represent symbol entropy; 4-stage Multi-Stage Vector Quanitzation of LSF parameters at 24 bpf.

Stage No.	1	2	3	4
1	2.05	0.46	0.28	0.22
2	0.46	0.57	0.23	0.17
3	0.28	0.23	0.28	0.17
4	0.22	0.17	0.17	0.20

Table 2. Mutual information of MSVQ stages between successive frames in bits; 4-stage Multi-Stage Vector Quanitzation of LSF parameters at 24 bpf.

Table 1 presents the mutual information of different MSVQ stages for quantization of LSF parameters in the training database. Noticeable dependency is observed, particularly between the first stage and other stages of the MSVQ. Note that the diagonal terms represent the symbol entropy.

Table 2 presents the mutual information of different MSVQ stages across consecutive time instants, for quantization of LSF parameters in the training database. This indicates the interframe dependency of MSVQ output symbols. As seen noticeable dependency exists, and specially there is more than 2 bits of redundancy between the first MSVQ stages over time. This is due to the high time correlation of LSF parameters.

4.2. Numerical Results

We evaluate four decoders for reconstruction of LSF parameters encoded with the described MSVQ.

The decoder MS0 provides the baseline for comparisons and is given by

$$\hat{\mathbf{x}}_n = \sum_{k=1}^{\eta} \sum_{I_n^{(k)}} \tilde{\mathbf{Y}}_n^{(k)} P(J_n^{(k)}|I_n^{(k)}) \qquad (10)$$

This decoder is often refered to as the *Basic MMSE* decoder and does not exploit any residual redundancy. A variant of this decoder, is suggested for decoding of MSVQ encoded signals in [17].

The decoder MS1 is given in equation (6), and exploits both intraframe and interframe dependencies. Examining the intraframe dependencies of LSF parameters, provided in Table 1, we selected the values of k' and k'' according to Table 3 to incorporate a high level of residual redundancy.

In some applications, the successive refinement feature of the MSVQ is utilized in the source coder design. In such cases, the decoder should also have the same characteristic, i.e., decoding each

---

[1] The speech databases used in this work are provided by Nortel Networks.

k	k'	k''	k	k'	k''
1	2	3	3	1	2
2	1	3	4	1	3

Table 3. The values of parameters k' and k'' in equation (6) for reconstruction of LSF parameters using decoder MS1.

Channel SNR (dB)	BER	MS0	MS1	MS2	MS3
1.00	0.0560	3.68	2.57	2.65	3.27
2.00	0.0370	3.05	2.15	2.22	2.68
3.00	0.0220	2.41	1.78	1.82	2.13
4.00	0.0120	1.85	1.47	1.50	1.67
5.00	0.0059	1.45	1.26	1.27	1.35
6.00	0.0023	1.20	1.14	1.14	1.16
7.00	0.0008	1.08	1.08	1.08	1.08

Table 4. Average spectral distortion (dB) of the test LSF database reconstructed using four decoding schemes for transmission over an AWGN channel with soft outputs and BPSK modulation.

stage of the MSVQ must be independent of future stages. This is not the case in MS1. Therefore, we present decoder MS2 as a progressive version of the decoder MS1. That is, in reconstruction of each MSVQ stage only the intraframe dependency with prior stages are exploited. Although, this feature of the MSVQ is not utilized in MELP, however, the numerical results provide an insight into the incurred level of performance degradation with respect to MS1, to maintain the progressive feature of the decoder.

The decoder MS3 is also a variant of the decoder MS1, which exploits only the intraframe dependencies, but ignores the interframe dependencies. This decoder is given by

$$\hat{\mathbf{x}}_n = \sum_{k=1}^{\eta} \sum_{I_n^{(k)}} \tilde{\mathbf{Y}}_n^{(k)} P(I_n^{(k)} | J_n^{(k)}).P(J_n^{(k')} | I_n^{(k)}).P(J_n^{(k'')} | I_n^{(k)}). \tag{11}$$

Table 4 presents the performance of the decoders MS0, MS1, MS2 and MS3 for reconstruction of the LSF test database encoded with the 24 bpf MSVQ. The MSVQ output bits are transmitted using a BPSK modulation, over an AWGN channel, which provides soft outputs to the MSVQ decoder. As seen from Table 4, the decoders that exploit the residual redundancy achieve noticeable gains, in comparison with the basic MMSE decoder MS0. Specifically, using decoder MS1 the spectral distortion in the reconstructed signal is reduced by more than 1dB, for very noisy channels. This is followed, rather closely, by the progressive decoder MS2. Examining the performance of the decoder MS3, it is observed that about 40% of the gain of MS1 is due to the intraframe residual redundancy.

5. REFERENCES

[1] 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.

[2] 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.

[3] 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, May 2001.

[4] R. Link and S. Kallel, "Optimal use of Markov models for DPCM picture transmission over noisy channels." *IEEE Trans. Image Proc.,* vol. 48, No. 10, pp. 1702-1711, 2000.

[5] M. Park and D. J. Miller, "Improved Image decoding over noisy channels using mmse estimation and a Markov mesh," *IEEE Trans. Image Proc.,* vol. 8, No. 6, pp. 863-867, 1999.

[6] 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.

[7] T. Fingscheidt, T. Hindelang, R. V. Cox, and N. Seshadri, "Joint source-channel (de-)coding for mobile communications," *IEEE Trans. Commun.,* vol. 50, No. 2, pp. 200-212, 2002.

[8] 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.

[9] 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, 1996.

[10] A. V. McCree and T. P. Barnwell, III, "A 2.4 kbits/s MELP coder candidate for the new U.S. federal standard," *Proc. Int. Conf. Acoust.,Speech, Signal Process.,* pp. 200-203, 1996.

[11] C. F. Barnes, S. A. Rizvi, and N. M. Nasrabadi, "Advances in residual vector quantization: a review," *IEEE Trans. Image Proc.,* vol. 5, No. 2, pp. 226-262, 1996.

[12] H. Khalil and K. Rose, "Asymptotic closed-loop design of predictive multi-stage vector quantizers," *Proc. Int. Conf. Image Proc.,* vol. 2 , pp. 199 -202.

[13] W. F. LeBlanc, B. Bhattacharya, S. A. Mahmoud and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Trans. Speech and Audio Process.*, vol. 1, no. 4, pp. 373-385, 1993.

[14] D. H. Lee and D. L. Neuhoff, "Asymptotic distribution of the errors in scalar and vector quantizers," *IEEE Trans. Inform. Theory,* vol. 42, pp. 446-460, Mar.1996.

[15] A. B. Sripad and D. L. Snyder, "A necessary and suffcient condition for quantization errors to be uniform and white," *IEEE Trans. Acoustics, Speech, and Signal Proc.,* vol.25, No.5, pp. 442- 448, 1977.

[16] B. Widrow, I. Kollr, and M.- C. Liu "Statistical theory of quantization," *IEEE Trans. on Instrumentation and Measurement,* 45(2), pp. 353- 361, 1996.

[17] S. Gadkari and K. Rose, "Unequally protected multistage vector quantization for time-varying CDMA channels," *IEEE Trans. Commun.*, vol. 49, no. 6 , pp. 1045-1054, 2001.