MSVQ DESIGN FOR PACKET NETWORKS WITH APPLICATION TO LSF QUANTIZATION

Hosam Khalil and Kenneth Rose

Signal Compression Laboratory
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106, USA
hkhalil@scl.ece.ucsb.edu rose@ece.ucsb.edu

ABSTRACT

The design of a multi-stage vector quantizer (MSVQ) based source-channel coding system is optimized for packet networks. Resilience to packet loss is further enhanced by a proposed interleaving approach that ensures that a single lost packet only eliminates a subset of the vector stages. The design is optimized while taking into account compression efficiency, packet loss rate, and the interleaving technique in use. The new source-channel-optimized MSVQ is tested on memoryless line spectral frequency (LSF) parameter quantization in speech coders. A source-channel-optimized MSVQ is shown to yield a gain of up to 2.0 dB in SNR, for coding the LSFs, over traditional MSVQ and to substantially enhance the robustness of packetized speech transmission. Although the formulation is given in the context of packet networks, the work is directly extendible to the broader category of erasure channels.

1. INTRODUCTION

Packet networks have dramatically gained in importance and popularity in recent years, especially due to the widespread use of the Internet. Naturally, much research effort is currently focused on robustness to the type of errors that are encountered in such channels, namely, packet loss due to congestion and delays. Packet networks, hence, represent a special case of erasure channels. This paper is concerned with the design of source-channel coding systems for packet networks in particular, and erasure channels in general. Another potential application is in wireless communications over deep fading channels which may also be viewed as erasure channels.

A known approach to combat packet loss is the design of diversity systems which employ multiple description coding. This direction of research was stimulated by Vaishampayan's work on scalar quantizers [1] and was later pursued in the context of transform coding [2]. The basic idea is to encode and transmit more than one description of the source over different channels or packets. In the event of packet loss or channel failure, the decoder reconstructs the data from the remaining received descriptions. The coder is thus designed such that there is sufficient redundancy in the descriptions. Other notable techniques to mitigate problems of data loss exist in the literature. See, for example, a review of robust audio coding in [3].

Multi-stage vector quantization (MSVQ) is widely used in compression of audio signals. MSVQ decomposes the source vector into the sum of code vectors, one per stage. Historically, MSVQ was conceived as a sequential quantization operation where each stage simply quantizes the residual of the previous stage. More recently, the greedy nature of simple sequential encoding was recognized, and efficient techniques were proposed to seek better approximation of the source vector as combination of stage-vectors [4]. In [5], an unequally protected MSVQ design was proposed where the receiver estimates the channel conditions and decodes as many stages of the quantized signal as can be reliably decoded. In that work, the multi-stage coder is viewed as a tool for successive refinement, and thus to decode stage n, all stages up to $n-1$ need to be correctly decoded. In the current work, we directly optimize MSVQ for general information loss patterns.

Existing methods in the speech coding literature offer certain resilience to information loss by separate encoding of partial information, e.g., TwinVQ [6] and conjugate structure CELP [7]. However, the resilience to information loss is not directly optimized.

This paper is organized as follows. In Section 2, we introduce basic means to exploit the robustness potential of MSVQ, via appropriate interleaving techniques. In Section 3, we propose a design technique to optimize the MSVQ for the given packet loss statistics. We then present simulation results and conclusions in Sections 4 and 5, respectively.

2. INTERLEAVING FOR ROBUST MSVQ

In this section we demonstrate the importance of appropriate interleaving to the robustness of MSVQ, and propose such an interleaving scheme. An L-stage MSVQ quantizes an input vector by searching for a set of indices $(a_0, a_1, a_2, \ldots, a_{L-1})$ pertaining to the (hopefully) best choice of stage-vectors for its representation. The process of finding a good combination of representative stage-vectors typically

This work is supported in part by the NSF under grant MIP-9707764, the University of California MICRO Program, Cisco Systems, Inc., Conexant Systems, Inc., Dialogic Corp., Fujitsu Laboratories of America, Inc., General Electric Co., Hughes Network Systems, Lernout & Hauspie Speech Products, Lockheed Martin, Lucent Technologies, Inc., Qualcomm, Inc., and Texas Instruments, Inc.

Table 1: Performance comparison of LSF reconstruction using source-channel-optimized MSVQ-indices and traditional source-optimized MSVQ. Shown is the average MSE at conditions of partial received information according to shown transmission error patterns. Also shown is the MSE when the previous reconstructed LSF is used as concealment in place of the missing LSF.

Transmission Vector	Source-Optimized MSVQ	Source-Channel-Optimized MSVQ
	MSE	MSE
11111	0.30	0.37
11110	0.69	0.82
11101	1.07	1.10
11100	1.39	1.53
11011	1.88	1.91
11010	2.22	1.98
11001	2.53	2.27
11000	2.80	2.78
10111	3.70	1.77
10110	4.06	2.35
10101	4.39	2.66
10100	4.67	3.22
10011	5.03	3.07
10010	5.33	3.73
10001	5.59	4.05
10000	5.82	4.69
Last Reconstructed	8.76	8.76

involves an "M-search" [4]. Instead of greedingly selecting the best vector at the current stage, the decision is postponed, and M "survivors" are temporarily stored. In other words, the search process puts more emphasis not on finding the best stage-vector at each stage, but the best total combination of stage-vectors. This fact, coupled with the observation that, at high dimensionality, individual stage-vectors exhibit a high degree of mutual orthogonality, suggests that it is worthwhile to decode and use a stage-vector even if some preceding indices are missing. This can be demonstrated statistically.

MSVQ is widely applied to coding of line spectral frequency (LSF) parameters, which constitute a major portion of the bitstream. An experiment was conducted to test the decoding properties of MSVQ at different index loss patterns and the results are shown in Table 1. Results are averaged over a test set of 6,000 LSF vectors. For a given L-stage MSVQ, the pattern of index losses can be described using an index transmission vector $T = (T_0, T_1, \ldots, T_{L-1})$ where

$$T_i = \begin{cases} 1 & \text{if index } i \text{ is correctly received} \\ 0 & \text{if index } i \text{ is lost.} \end{cases} \quad (1)$$

In the first column, we show the index transmission vectors. The second and third columns list for traditional MSVQ

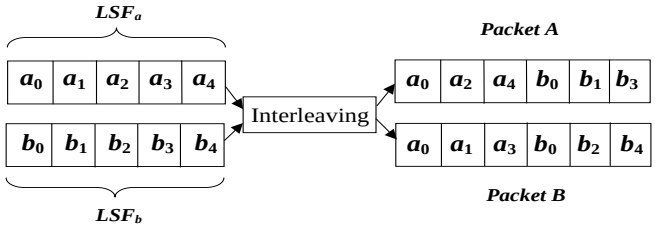

Figure 1: An example of interleaving of LSF parameters between two data packets. Note that a_0 and b_0 are transmitted on both packets.

design and source-channel optimized design, respectively, the mean squared error (MSE) results for LSF reconstruction subject to the corresponding index transmission vector, averaged over the test set. The Table demonstrates the improvements of channel-optimized MSVQ over traditional MSVQ. In the next section, we will explain the design steps needed to achieve this improvement. For reference, we also included in the last row the average MSE when the last correctly reconstructed LSF is repeated for concealment – the most widely adopted recovery technique when speech information is lost [3].

From Table 1, we can see that, on average, using all received indices for the reconstruction is preferable to repeating the previous LSF (as if the whole LSF is lost). Note that all transmission patterns indicate successful transmission of the first stage-index. The requirement that a_0 always be received is due to the fact that a_0 contains the most significant information, and the loss of a_0 often renders the remaining indices useless. For this reason, in some audio standards, a_0 is more heavily protected using FEC codes.

We propose a framework for encoding the LSFs into data packets such that the decoding is more robust to channel losses. Table 1 strongly suggests that it is better to have a loss distributed over several LSFs than to have a complete loss of an LSF, which would then require concealment as mentioned above. An interleaving method is the natural solution to this problem. See Fig. 1 for an example of interleaving of two LSFs over two packets. Notice that a_0 and b_0 are transmitted on both packets. Although this redundancy increases the bit rate, it was found to contribute significantly to robustness.

Using the interleaving scheme of Fig. 1, and given packet loss rate of p, we have four different transmission vectors: 00000, 10101, 11010, 11111, with probabilities p^2, $p(1-p)$, $(1-p)p$, and $(1-p)(1-p)$, respectively. The transmission vector 00000 occurs when both packets are lost, and in this case the last received LSF is used for concealment.

3. SOURCE-CHANNEL-OPTIMIZED MSVQ

In this section, we propose an optimization algorithm to further strengthen MSVQ for use on packet network channels under interleaving schemes such as the one described in Section 2.

The traditional approach to the design of MSVQ is an iterative decent algorithm alternating between two optimization steps. Given the current quantizer codebooks, the

training set is encoded, and statistics for the input vectors for each stage quantizer and its decision are accumulated. Next, codebook entries are recalculated from the current encoding partition statistics. These two steps guarantee monotone decrease in reconstruction error, and the procedure converges to an at least locally optimal solution.

The aim of the new algorithm is to extend the design algorithm to handle packet loss. In fact, we average over all possible transmission vectors with respect to probabilities p_T, where $T \in \mathcal{T}$, the set of possible transmission vectors. For an L-stage MSVQ, an input vector (of dimension n) $X^n \in \mathcal{X}^n$ is encoded into L indices $(i_0, i_1, \ldots, i_{L-1})$ that are concatenated into one overall codeword i. The set $\mathcal{I}$ is the set of all possible codewords such that $i \in \mathcal{I}$. It is assumed that the channel codeword is the same as the source codeword. Thus, we denote the encoder and decoder by $\alpha : \mathcal{X}^n \to \mathcal{I}$ and $\beta : \mathcal{I}' \to \mathcal{Y}^n$, respectively. The function β, thus, maps codeword i' to a reproduction vector $Y^n \in \mathcal{Y}^n$.

Encoder Optimization

An optimal encoder α^* maps each source vector X^n to the codeword that would minimize the expected distortion. If transmitted codeword i is subject to channel losses in the pattern of the transmission vector T, then the received codeword is denoted by $f(i,T)$. Lost indices within i are assumed to result in zero-valued stage-codevectors. Thus, the optimal $\alpha^*(X^n)$ is given by

$$\alpha^*(X^n) = \arg\min_{i \in \mathcal{I}} \{ \sum_{T \in \mathcal{T}} p_T d(X^n, \beta(f(i,T))) \} \qquad (2)$$

where $d(\cdot,\cdot)$ is the distortion measure. Note that for simplicity of description we have made abstraction of the stage-wise encoding search.

Decoder Optimization

Given α, we find the optimal decoder β^* that minimizes the expected distortion given the received channel codeword index $i' \in \mathcal{I}'$:

$$\beta^*(i') = \arg\min_{Y^n \in \mathcal{Y}^n} \{ E_{X^n}[d(X^n, Y^n) \mid f(\alpha(X^n), T) = i'] \} \qquad (3)$$

For a codeword i' that has been subjected to losses, there exists a set $\mathcal{M} \subset \mathcal{I}$ such that $m \in \mathcal{M}$ after losses T will lead to the same codeword as the received codeword i', i.e., $f(m,T) = i'$. Thus, the expected value that we minimize may be written as

$$E_{X^n}[d(X^n, Y^n) \mid f(\alpha(X^n), T) = i'] = \sum_{i \in \mathcal{M}} E_{X^n}[d(X^n, Y^n) \mid \alpha(X^n) = i] \frac{P(\alpha(X^n) = i)}{P(f(\alpha(X^n), T) = i')} \qquad (4)$$

An explicit solution for $\beta^*(i')$ for the case of squared error distortion is given by

$$\beta^*(i') = \sum_{i \in M} E_{X^n}[X^n \mid \alpha(X^n) = i] \frac{P(\alpha(X^n) = i)}{P(f(\alpha(X^n), T) = i')} \qquad (5)$$

The iterative algorithm, which alternates between (2) and (5), must be initialized with some MSVQ. A reasonable choice of initialization is with an MSVQ that was optimized assuming a lossless channel (i.e., a source-optimized MSVQ). The actual algorithm we use here is implemented using a selective splitting procedure [8].

4. SIMULATION RESULTS

A complete robust speech coder would require error-resilient techniques for all the transmitted components such as gain, pitch, excitation, and LSF parameters. In this paper, we provide preliminary results that concentrate on the LSF portion of the coder which is quantized by MSVQ. The MSVQ parameters are encoded into the bitstream and constitute a considerable portion of the allocated bandwidth. For example, in MELP, LSF parameters use 25 bits of the allocated 54 bits per frame of speech.

For the experiments, we assume a packet-based communication channel. Known interleaving techniques, such as reviewed in [3], or proposed in the Internet RTP (real time protocol) payload format for interleaved media, perform interleaving on the scale of whole speech frames, which we call frame-interleaving. In this work, we propose an interleaving scheme that operates on the much finer level of MSVQ-indices, which we refer to as MSVQ-index interleaving. The MSVQ-index interleaving scheme of Fig. 1 will be used.

In the first experiment, we compare frame-interleaving with MSVQ-index interleaving. The MSVQ-index interleaving is used without optimization to demonstrate the power of the interleaving scheme alone. A MELP coder is used and only the LSF portion of the bitstream is subject to packet loss. In MELP, 25 bits are devoted to LSF quantization. We use a slightly different structure for the quantizer from the one used in the standard. A 5-stage MSVQ is used, with each stage equally allocated 5 bits. Since the interleaving procedure used (Fig. 1) assumes sending the first stage index twice, the total bit rate used here is 30 bits/LSF. For the case of frame-interleaving, we test two cases: Case A, frame-interleaving uses 5 stages , i.e. 25 bits/LSF, while in case B, frame-interleaving uses 6 stages, i.e. 30 bits/LSF. We use case B to match the bit rate of the MSVQ-index interleaving coder. See Table 2 for the results of an informal listening test involving 12 listeners and a representative set of MIRS sentences. Note that the proposed MSVQ-index interleaving method is significantly preferred over the traditional frame-interleaving method even without source-channel optimization. When the bit rate used is exactly the same (case B), the "non-optimized" MSVQ-index interleaving method still outperforms the frame-interleaving by achieving about 78% preference.

In the second experiment, we further improve the performance of the MSVQ-index interleaving method by optimizing the system for packet losses. A source-channel-optimized MSVQ can thus be optimized using the design of Section 3, and the given rate of transmission errors as described in Section 2. Here, we compare the source-optimized MSVQ-index interleaving with that of the more powerful source-channel-optimized MSVQ-index interleaving. Note that both methods use our proposed interleaving scheme. The objective is to evaluate the added benefit of source-channel design. In Table 3, we present informal speech listening tests at five different packet loss probabilities. It can be seen that a source-channel-optimized MSVQ signif-

Table 2: Preferences in an informal listening test involving 12 listeners. Shown is a comparison between traditional frame-interleaving and the proposed MSVQ-index interleaving. Frame-interleaving uses 25 bits in case A, and 30 bits in case B. MSVQ-index-interleaving uses 30 bits/LSF in both cases. Values shown in percent.

	Traditional Frame-Interleaving	MSVQ-Index-Interleaving	No Preference
Case A	8.33	88.33	3.34
Case B	15.00	78.33	6.67

Table 3: Preferences in an informal listening test involving 12 listeners. Source-optimized MSVQ versus proposed source-channel-optimized MSVQ. Values shown in percent.

	Source-Optimized MSVQ	Source-Channel-Optimized MSVQ	No Preference
$p = 0.1$	16.67	54.17	29.16
$p = 0.15$	8.33	75.00	16.67
$p = 0.2$	12.50	79.17	8.33
$p = 0.3$	12.50	62.50	25.00
$p = 0.5$	16.67	66.67	16.66
Average	**13.33**	**67.33**	**19.17**

icantly outperforms a source-optimized MSVQ and indeed further improves the speech quality of packetized speech. For an objective comparison, we show in Fig. 2 the mean squared error of the different source-channel coding strategies. Here, it can be seen that the source-channel optimized MSVQ outperforms all other strategies. A gain in SNR of up to 2.0 dB can be obtained over the traditional frame-interleaving method. Also, source-channel optimization outperforms the source optimization by up to 0.87 dB. A comparison in terms of spectral distortion was found to be misleading. In this case the proposed source-channel optimized method outperforms the source optimized method by up to 0.3 dB in spectral distortion, but the comparison between frame-interleaving and MSVQ-index interleaving was not clear cut. Since the listening tests in Table 2 show clear and substantial preference of the proposed method, we conclude that the type of sound artifacts due to concealment is not well captured by spectral distortion.

5. CONCLUSION

We proposed a new interleaving scheme which is suitable for robust speech transmission over lossy packet networks. Further, we developed an optimization algorithm for MSVQ design for packet-switched losses. Both subjective and objective comparisons prove the merit of the proposed approach.

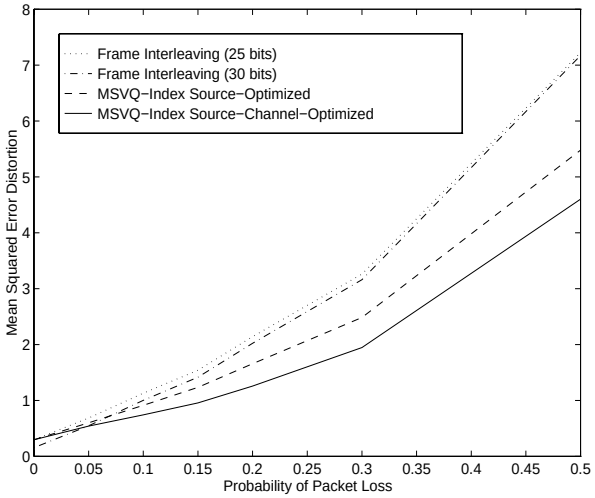

Figure 2: MSE performance comparison of the following: (i) Traditional frame-interleaving using 5 MSVQ stages (25 bits), (ii) Traditional frame-interleaving using 6 MSVQ stages (30 bits), (iii) Source optimized MSVQ-index interleaving using 5 stages packetized using 6 stages (30 bits), and (iv) Proposed source-channel optimized MSVQ-index interleaving using 5 stages packeted into 6 stages (30 bits).

6. REFERENCES

[1] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Information Theory,* vol. 39, no. 3, pp. 821-834, May 1993.

[2] Y. Wang, M. T. Orchard, and A. R. Reibman, "Optimal pairwise correlating transforms for multiple description coding," *ICIP'98,* Chicago, IL, vol. 1, pp. 679-683, Oct. 1998.

[3] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Network,* vol. 12, no. 5, pp.40-48, Sept./Oct. 1998.

[4] W. P. 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 Proc.* vol. 1, no. 4, pp. 373-385, Oct. 1993.

[5] S. Gadkari and K. Rose, "Unequally protected multistage vector quantization for time-varying channels," *ICC'98,* Atlanta, GA, vol. 2, pp. 786-90, June, 1998.

[6] N. Iwakami, T. Moriya, and S. Miki, "High-quality audio coding at less than 64 kbit/s by using TwinVQ," *Proc. ICASSP'95,* pp. 937-940, 1995.

[7] A. Kataoka, T. Moriya, and S. Hayashi, "An 8-kb/s conjugate structure CELP (CS-CELP) speech coder," *IEEE Trans. Speech and Audio Proc.,* vol. 4, no. 6, pp. 401-411, Nov. 1996.

[8] H. Khalil and K. Rose, "A selective splitting approach to entropy-constrained single/multi-stage vector quantization design," *Image and Video Communications and Processing 2000,* San Jose, CA, Jan. 2000.