Robust Jointly Optimized Multistage Vector Quantization for Speech Coding

Venkatesh Krishnan and David V. Anderson

School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA 30332-0250, USA

10.21437/Eurospeech.2003-360

Abstract

In this paper, a novel channel–optimized multistage vector quantization (COMSVQ) codec is presented in which the stage codebooks are jointly designed. The proposed codec uses a signal source and channel–dependent distortion measure to encode line spectral frequencies derived from segments of a speech signal. Simulation results are provided to demonstrate the consistent reduction in the spectral distortion obtained using the proposed codec as compared to the conventional sequentially–designed channel–matched multistage vector quantizer.

1. Introduction

Vector Quantization (VQ) is extensively used in several commercially popular speech, image, and video coding standards. In particular, the use of vector quantizers in source–system model based coding of speech signals has enabled high quality speech coding at very low bit rates.

The direct use of VQ as a practical signal coding technique in a communication system suffers from the following two limitations: (1) complexity of implementation for large vector dimensions and codebook sizes, and (2) sensitivity to errors in the received indices due to noise in the communication channel. Several structurally constrained vector quantizers, such as multistage vector quantizers (MSVQ) [1], reduce the complexity of implementation for a penalty in the reconstruction fidelity. The effect of channel errors can be mitigated by designing vector quantizers that are optimized for both the source and the channel characteristics.

A channel–matched MSVQ (CMMSVQ) design has been proposed in [2] in which the MSVQ codebooks are trained stage–by–stage by minimizing a source and channel–dependent distortion measure. Such a sequential design is suboptimal since, while designing a given stage, it assumes that the subsequent stages are populated by zero vectors.

In this contribution, we develop and present a channel–optimized multistage VQ (COMSVQ) codec in which the stage codebooks are jointly designed by minimizing a source and channel–dependent distortion measure. Each stage encoder of the proposed jointly–designed COMSVQ codec accounts for the effect of channel errors on the indices generated by all the stage encoders. The design and operation of the proposed codec is described in Section 2. In Section 3, we describe the application of the jointly–designed COMSVQ to encode the line spectral frequencies (LSFs) [3] obtained from windowed segments of a speech signal. Simulation results are provided to demonstrate the improved performance of the proposed COMSVQ, compared to that of the conventional sequentially–designed CMMSVQ.

2. Channel–Optimized MSVQ of LP Parameters

Consider a database, $\mathbf{V}$, of n–dimensional LSF vectors. Let ζ be a set of integers. An LSF VQ encoder can be thought of as a mapping $\mathbf{Q} : \mathbf{x} \rightarrow i$, that maps the vector $\mathbf{x} \in \mathbf{V}$ to an integer $i \in \zeta$. Typically, i is selected to be the index of the codevector $\mathbf{C}_i$ in a codebook $\mathcal{C}$ that minimizes a predetermined distortion measure $D(\mathbf{x}; \mathbf{C}_i)$. If the codebook has N codevectors $\{\mathbf{C}_i, i = 0, ...N-1\}$, then $\zeta \triangleq \{0, 1, ..., N-1\}$. The VQ decoder $\widehat{\mathbf{Q}} : i \rightarrow \mathbf{C}_i$ maps the index i back to the codevector $\mathbf{C}_i$. Thus the VQ codec quantizes $\mathbf{x}$ to $\widehat{\mathbf{Q}}\big(\mathbf{Q}(\mathbf{x})\big) = \mathbf{C}_i$.

An MSVQ is a structurally constrained VQ in which $\mathbf{x}$ is encoded by K successive VQ encoders $(\mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_K)$. In a source optimized MSVQ (SOMSVQ), the k^{th} stage maps the residual input, $\mathbf{x}_k = \mathbf{x} - \sum_{m=1}^{k-1} \mathbf{C}^{(m)}_{i_m}$, onto the index i_k of the codevector $\mathbf{C}^{(k)}_{i_k}$ in the stage codebook $\mathcal{C}^{(k)}$. Thus, $\mathbf{x}$ is quantized to the sum of appropriate codevectors selected from all the K stages. However, in the presence of channel errors, the set of indices generated by the encoders, $I = \{i_1, i_2, ...i_K\}$, is not the same as that received by the corresponding decoders, denoted $J = \{j_1, j_2, ...j_K\}$. This results in a distortion that can be accounted for by including the channel characteristics in the distortion measure used in designing the codebooks of the COMSVQ. In the following subsection, we propose a joint design algorithm, instead of conventional suboptimal sequential design, for training the COMSVQ codebooks.

2.1. Joint COMSVQ Codebook Design

Transmission of the index set I generated by the MSVQ over a noisy channel, results in a distorted index set J at the decoder. Thus, the transition probability $P(J|I)$ may be used to characterize the channel's statistical properties. Since the codec, in the presence of channel errors, reconstructs $\mathbf{x}$ as $\sum_{m=1}^{K} \mathbf{C}_{j_m}^{(m)}$, the expected value of the distortion suffered by $\mathbf{x}$ over all possible transitions of I is given by

$$D(\mathbf{x}; \mathbf{C}_{i_1}^{(1)}, \mathbf{C}_{i_2}^{(2)}, ..., \mathbf{C}_{i_K}^{(K)}) = \sum_J P(J|I) d\big(\mathbf{x}, \sum_{m=1}^{K} \mathbf{C}_{j_m}^{(m)}\big) \quad (1)$$

In general, $d(\mathbf{x}, \mathbf{y})$ may defined as a weighted Euclidian distance between $\mathbf{x}$ and $\mathbf{y}$ with a positive diagonal weighing matrix $\mathbf{W}$. If the source has an n-fold output probability distribution function $f_{\mathbf{x}}(\mathbf{x})$, then the average distortion $\mathbf{D}$ is given by

$$\mathbf{D} = \sum_I \int_{\mathbf{V}_I} D(\mathbf{x}; \mathbf{C}_{i_1}^{(1)}, \mathbf{C}_{i_2}^{(2)}, ..., \mathbf{C}_{i_K}^{(K)}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}, \quad (2)$$

where the partition $\mathbf{V}_I$ of the signal space $\mathbf{V}$ is given by

$$\begin{aligned} \mathbf{V}_I &= \Big\{\mathbf{x} : D(\mathbf{x}; \mathbf{C}_{i_1}^{(1)}, \mathbf{C}_{i_2}^{(2)}, ..., \mathbf{C}_{i_K}^{(K)}) \\ &\leq D(\mathbf{x}; \mathbf{C}_{l_1}^{(1)}, \mathbf{C}_{l_2}^{(2)}, ..., \mathbf{C}_{l_K}^{(K)}) \quad \forall l_k \in \zeta \Big\}. \end{aligned} \quad (3)$$

The K stage encoders of COMSVQ are jointly designed by using a suitably modified version of the generalized Lloyd's algorithm [1] on a training database of vectors as described below. Each iteration of the design algorithm consists of the following two basic steps: (1) determining the partition $\mathbf{V}_I$ (Eq. 3) of the signal space $\mathbf{V}$ for all I, and (2) updating the codevectors, $\mathbf{C}_{j_k}^{(k)}$, of each stage codebook (i.e, $k = 1, 2, ...K$) corresponding to the partitions obtained in step (1), so that the average distortion of the entire codec, $\mathbf{D}$ is minimized. This can be achieved by setting

$$\nabla_{\mathbf{C}_{j_k}^{(k)}} \mathbf{D} = 0. \quad (4)$$

The explicit solution of Eq. 4 can be readily obtained by substituting for $\mathbf{D}$ from Eq. 2.

$$\nabla_{\mathbf{C}_{j_k}^{(k)}} \mathbf{D} = \sum_I \int_{\mathbf{V}_I} \Big[\sum_{J-\{j_k\}} P(J|I) 2\mathbf{W}\big(\mathbf{x} - \mathbf{C}_{j_k}^{(k)} - \sum_{\substack{m=1 \\ m \neq k}}^{K} \mathbf{C}_{j_m}^{(m)}\big) \Big] f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = 0 \quad (5)$$

Thus, the codevector $\mathbf{C}_{j_k}^{k}$ is updated according to,

$$\mathbf{C}_{j_k}^{(k)} = \Big[\sum_I \int_{\mathbf{V}_I} \mathbf{W} \Big\{ \sum_{J-\{j_k\}} P(J|I) \Big\} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \Big]^{-1} \Big[\sum_I \int_{\mathbf{V}_I} \mathbf{W}\, \hat{\mathbf{x}}_k\, f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \Big], \quad (6)$$

where the expected value of the residual input to the k^{th} stage, $\hat{\mathbf{x}}_k$ is given by

$$\hat{\mathbf{x}}_k = \sum_{J-\{j_k\}} \Big[P(J|I) \Big(\mathbf{x} - \sum_{\substack{m=1 \\ m \neq k}}^{K} \mathbf{C}_{j_m}^{(m)} \Big) \Big]. \quad (7)$$

The joint codebook design algorithm is summarized below. Since the partitions and the stage codebooks are modified every iteration, the iteration number t is explicitly included as (t).

1. *Initialization*: The iteration number, t is set to 1. The initial codebooks are set to $\mathcal{C}(0) = \{\mathcal{C}^{(1)}(0), ..., \mathcal{C}^{(K)}(0)\}$. k is initialized to 1.

2. *Partition of the training set*: Using the latest set of codebooks, $\mathcal{C}(t-1)$, all partitions of the database are determined (Eq. 3). This step associates a set of indices I with every training vector $\mathbf{x}$. The M–candidate search procedure with an appropriate value of M is employed in evaluating Eq. 3.

3. *Termination criterion check*: The average distortion function at iteration t, $\mathbf{D}(t)$, is evaluated. The training is terminated if $|\mathbf{D}(t) - \mathbf{D}(t-1)|/\mathbf{D}(t)$, drops below a predetermined threshold value, δ.

4. *Codebook update*: Each codevector in the k^{th} stage, $(j_k = 1, ..., N)$, is updated according to Eq. 6.

5. *Repeat*: Both t and k are incremented by 1. If $k = K$, k is re–initialized to 1. The algorithm loops back to step 2.

2.2. COMSVQ Codec Operation

In this section, we describe the operation of the COMSVQ encoder/decoder pair whose stage codebooks are designed jointly as described in Section 2.1. To encode a vector $\mathbf{x}$, an optimal (full) search requires considering all possible combinations of codevectors from all stage codebooks and selecting the one that minimizes Eq. 1. For the typical codebook sizes encountered in coding LSFs, a full search is computationally prohibitive. The codec presented in this paper employs a suboptimal M-candidate search [4] of the codebooks for a marginal decrease in the reconstruction quality. In the M-candidate search procedure, M codevectors that give the lowest distortion, $D(\mathbf{x}; \mathbf{C}_{i_1}^{(1)})$, are selected from the first stage codebook. Then, the second stage codebook is searched M times, once for each vector chosen from the first stage. The M paths that give the lowest overall distortion, $D(\mathbf{x}; \mathbf{C}_{i_1}^{(1)} + \mathbf{C}_{i_2}^{(2)})$, are selected. This procedure is repeated for all K stages. It must be noted that

Bit Error Rate	SO-MSVQ	CM-MSVQ			JD-CO-MSVQ		
		M=1	M=2	M=4	M=1	M=2	M=4
0	1.3785	1.3785	1.3598	1.3134	1.2154	1.2105	1.1905
0.0001	1.4871	1.462	1.408	1.3789	1.3519	1.2884	1.2547
0.001	1.5749	1.5166	1.4596	1.423	1.4219	1.362	1.3175
0.005	2.0034	1.7803	1.7244	1.7187	1.6964	1.6225	1.6161
0.01	2.5439	2.0585	2.0305	2.0176	1.9572	1.929	1.9154

Table 1: Comparison of average SD for 5000 test vectors from TIMIT-Test database for three cases: (I) SOMSVQ, (II) Three Stage CMMSVQ, (III) Three Stage Jointly Designed-COMSVQ

each encoder stage of the jointly–designed COMSVQ accounts for the possible distortions suffered by the set of indices I. The proposed COMSVQ decoder receives the set of indices J and reconstructs $\mathbf{x}$ as the sum of the corresponding entries from the stage codebooks.

3. Jointly–Designed COMSVQ for LSF Quantization

In this section, we describe the implementation of the proposed jointly–designed COMSVQ codec, for $K = 3$, to encode the LSFs obtained from 10 ms frames of a speech signal using a Hamming window with 20% overlap between frames. To design the COMSVQ codebooks, 170,000 LSF vectors are obtained from speech records in the TIMIT training database. The perceptual weighing matrix $\mathbf{W}(\mathbf{x})$ given in [3] is used in the distortion function. For simplicity of formulation, we assume that the channel is binary symmetric with a known bit error rate, and that the indices suffer distortion independently. In other words,

$$P(J|I) = \prod_{m=1}^{K} P(j_m|i_m). \tag{8}$$

The joint COMSVQ codebook design algorithm is initialized with a set of codebooks trained on a portion of the database using the sequentially–designed CMMSVQ described in [2]. The codebooks for the multistage vector quantizer are designed as described in Section 2.1. Each stage codebook has 256 codevectors.

The performance of the proposed jointly–designed CO MSVQ is evaluated for 5000 test vectors derived from the TIMIT testing database in terms of the average SD of the reconstructed LSFs. The comparison between the average SD of the reconstructed LSFs obtained using two stage implementations of (1) SOMSVQ, (2) CMMSVQ described in [2], and (3) the proposed jointly–designed COMSVQ, is presented in Table 1. It is observed that the jointly–designed COMSVQ codec outperforms the sequentially–designed CMMSVQ by approximately 0.12 dB.

The presence of noise in the channel corrupts the reconstruction of the corresponding LP coefficients. The SD corresponding to the speech reconstructed from these "outlier" vectors is typically larger than 2 dB. The degradation in this reconstructed speech can often be heard as "clicks". The percentage of outliers with more than 2 dB of SD two cases: (I) Three Stage CM-MSVQ, (II) Three Stage Jointly Designed-CO-MSVQ, is shown in Figure 1.

It must be noted that although the CMMSVQ and the proposed jointly–designed COMSVQ differ in their codebook training algorithms, their encoders (and decoders) are identical. Thus, the encoding complexity of the jointly–designed COMSVQ is the same as that of the CMMSVQ for the same values of the parameter M of the M-candidate search.

4. Conclusions

In this contribution, we develop and present a novel channel–optimized multistage vector quantization codec whose codebooks are jointly designed to minimize a source and channel–dependent distortion measure. The proposed codec is used to encode LSFs obtained from speech signals. Simulation results demonstrate the superior performance of the proposed codec in terms of the average spectral distortion compared to the conventional

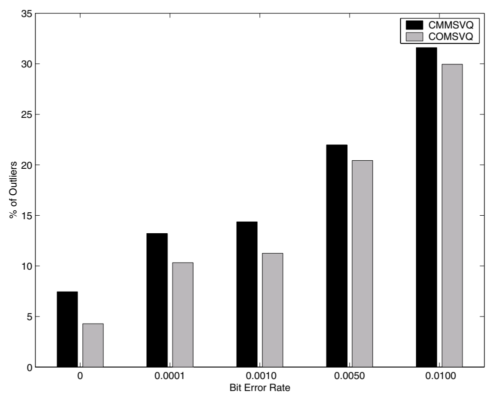

Figure 1: Percentage of outliers with more than 2 dB of SD for 5000 test vectors from TIMIT-Test database for two cases: (I) Three Stage CMMSVQ, (II) Three Stage Jointly Designed-COMSVQ

sequentially–designed channel–matched multistage vector quantizer. Moreover, the performance improvement is achieved without any additional bit requirements.

5. References

[1] Gersho A and Gray R.M, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1991.

[2] Phamdo N, Farvardin N, and Moriya T, "A unified approach to tree structured and multistage vector quantization for noisy channels," IEEE Transactions on Information Theory, vol 39, no 3, May 1993, pp. 835-850.

[3] Kang G and Fransen L, "Low–bitrate speech coders based on Line Spectral Frequencies," Naval Research Lab, report no. 8857, Nov 1984.

[4] Anderson J and Bodie J, "Least squares quantization in PCM ," IEEE Transactionson Information Theory, vol.IT-21, July 1975, pp. 379-387.