Eurospeech 2001 - Scandinavia

Joint Source-Channel Coding for Low Bit-Rate Coding of LSP Parameters

José L. Pérez-Córdoba, Antonio J. Rubio, Antonio M. Peinado, Angel de la Torre

Dpto. Electrónica y Tecnología de Computadores
Facultad de Ciencias. Universidad de Granada (Spain)
E-mail:jlpc/rubio/amp/atv@ugr.es

Abstract

This work presents a quantization technique for LSP parameters which results in a low bit-rate transmission while providing protection against channel errors. As a generalization of the so called Channel Optimized Vector Quantization (COVQ), Channel Optimized Matrix Quantization (COMQ) can remove intraframe and interframe LSP redundancy with the target of protecting the information sent through a channel in the presence of noise. Split COMQ is used in order to reduce storage requirements and complexity. Results show that Split COMQ gives better performance under certain error conditions and a lower bit rate transmission in all channel conditions compared to the reference quantization techniques.

1. Introduction

Efficient LPC quantization is a key question on speech coding. Many techniques have been proposed with the objective of reducing the bit-rate while maintaining a good quality of the corresponding synthesized speech. One approach to achieve these objectives is applying vector quantization (VQ). To reduce storage requirements and complexity Split VQ (SVQ) was proposed [4]. VQ removes the intraframe correlation of LSP parameters. One way to remove also interframe correlations is to apply matrix quantization (MQ) [8]. However, MQ introduces bigger delays, storage requirements and complexity. Split MQ (SMQ) was proposed [9] to reduce these bigger requirements.

The performance of these quantization techniques degrades with the presence of channel errors. To mitigate the effect of channel errors without increasing the bit rate, joint source-channel coding techniques are used. For example, Channel Optimized Vector Quantization (COVQ) [1] was proposed in the context of VQ and, as a generalization of COVQ, Channel Optimized Matrix Quantization (COMQ) [5] was proposed in the context of MQ.

In [5] the application of Split COMQ to LSP quantization is studied with the objective of removing the intraframe LSP redundancy maintaining the same number bits per frame as in the reference quantization technique. In the present work we extend the application of Split COMQ to remove the interframe redundancies of LSP parameters in addition to the intraframe ones, resulting in a coder with a low bit-rate transmission while protecting the parameters against channel errors.

This paper is organized as follows. In Section 2 COMQ is presented. Expressions for necessary optimal conditions are given. Section 3 discusses the application of Split COMQ to LSP quantization and summarizes characteristics of the coders. In Section 4 results on the performance evaluation of the COMQ are presented and the discussion about these results are reported. Finally, Section 5 contains conclusions.

2. COMQ Technique

In this Section we present the fundamentals of COMQ technique and the necessary optimal conditions are presented. To introduce COMQ technique, let us consider a real-valued independent and identically distributed (i.i.d.) source $\mathcal{X} = \{X_i\}_{i=1}^{\infty}$ with probability density function (pdf) $p(x)$. The source is to be encoded by means of a matrix quantizer (MQ) whose output is transmitted over a waveform channel. We consider a $k \times N$ matrix quantization process with M levels.

The COMQ system, as depicted in Figure 1, consists of a encoder mapping γ, a signal selection module and a decoder mapping β. The encoder $\gamma : \mathbb{R}^N \times \mathbb{R}^k \rightarrow \mathcal{I}$, where $\mathcal{I} = \{1, 2, \ldots, M\}$, is described in terms of a partition $\mathcal{S} = \{S_1, S_2, ..., S_M\}$ of $\mathbb{R}^N \times \mathbb{R}^k$ according to

$$\gamma(X) = i, \quad \text{if} \quad X \in S_i, \quad i \in \mathcal{I} \tag{1}$$

where $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N)$ is a typical source output matrix and $\boldsymbol{x}_i, i = 1, \ldots, N$ is a source vector. The signal selection module maps an index i to a signal $\boldsymbol{s}$ that is transmitted over the channel. The details of this module can be find in [5].

First, we consider that the channel is a AWGN channel. Therefore, the random channel output vector $\boldsymbol{r} = (r_1, r_2, \ldots, r_L)$ is related to the input vector $\boldsymbol{s} = (s_1, s_2, \ldots, s_L)$ through

$$r_l = s_l + n_l, \qquad l = 1, 2, ..., L \tag{2}$$

where L is the dimension of the signal constellation and n_l's are i.i.d. zero-mean Gaussian random variables with common variance $\sigma^2 = N_0/2$.

Finally, the decoder β makes an estimate $\widehat{X}$ of the source matrix based on the received vector (channel output) $\boldsymbol{r}$. We will restrict our study to hard-decision decoder, that is, the decoder β makes an estimate, $\hat{\boldsymbol{\imath}}$, of the index transmitted, i, represented by the signal $\boldsymbol{s}$, based on the received vector $\boldsymbol{r}$. Given

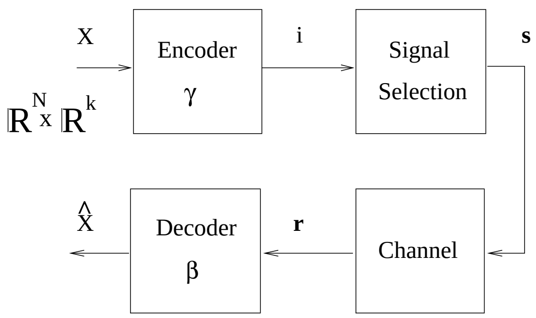


Figure 1: Block diagram of the COMQ system.

$\hat{\boldsymbol{\imath}}$, the estimate $\widehat{X}$ is selected from a finite reproduction alphabet (codebook) $\mathcal{C} = \{C_1, C_2, ..., C_M\}$ that described the decoder through

$$\beta(\hat{\boldsymbol{\imath}}) = \beta\left(\hat{\boldsymbol{\imath}}(\boldsymbol{r})\right) = C_{\hat{\boldsymbol{\imath}}}, \quad C_{\hat{\boldsymbol{\imath}}} \in \mathbb{R}^{\mathrm{N}} \times \mathbb{R}^{\mathrm{k}} \quad \hat{\boldsymbol{\imath}} \in \mathcal{I} \tag{3}$$

The performance of this system is generally measured by the average distortion per sample $\mathcal{D}(\mathcal{S}, \mathcal{C})$ and the encoding rate R. The average distortion is given by

$$\mathcal{D}(\mathcal{S}, \mathcal{C}) = \frac{1}{k} E\left[D\left(X, \beta\left(\hat{\boldsymbol{\imath}}(\boldsymbol{r})\right)\right)\right] \tag{4}$$

where $E[\cdot]$ means the expectation value and $D(X, Y)$ means the distortion measure used in the Generalized Linde-Buzo-Gray (GLBG) algorithm [8] defined by

$$D(X, Y) = \frac{1}{N} \sum_{n=1}^{N} d(\boldsymbol{x}_n, \boldsymbol{y}_n) \tag{5}$$

with $d(\boldsymbol{x}_n, \boldsymbol{y}_n) = \|\boldsymbol{x}_n - \boldsymbol{y}_n\|^2$. The encoding rate is given by

$$R = \frac{1}{kN} \log_2 M \ \text{ bits/sample} \tag{6}$$

The average distortion is a generalization to matrix quantization of the average distortion given in [1] for COVQ and it is given by

$$\mathcal{D}(\mathcal{S}, \mathcal{C}) = \frac{1}{k} \sum_{i=1}^{M} \int_{S_i} p(X) \left\{ \sum_{\hat{\boldsymbol{\imath}}=1}^{M} P(\hat{\boldsymbol{\imath}}|i) D(X, C_{\hat{\imath}}) \right\} \mathrm{d}X \tag{7}$$

where $p(X) = \prod_{n=1}^{N} p(\boldsymbol{x}_n) = \prod_{n=1}^{N} \prod_{i=1}^{k} p(x_{ni})$ is the kN-dimensional source pdf.

For a given source, a given channel, a fixed dimension k and N and a fixed codebook size M, we wish to minimize $\mathcal{D}(\mathcal{S}, \mathcal{C})$ by proper choice of $\mathcal{S}$ and $\mathcal{C}$.

2.1. Necessary Conditions and Algorithm

As in [1] and from (7) it becomes clear that for a fixed $\mathcal{C}$, the optimum partition $\mathcal{S}^* = \{S_1^*, S_2^*, ..., S_M^*\}$ is given by

$$\begin{aligned} S_i^* \quad = \quad & \left\{ X \ : \ \sum_{\hat{\boldsymbol{\imath}}=1}^{M} P(\hat{\boldsymbol{\imath}}|i) D(X, C_{\hat{\imath}}) \right. \\ & \left. \leq \ \sum_{\hat{\boldsymbol{\imath}}=1}^{M} P(\hat{\boldsymbol{\imath}}|l) D(X, C_{\hat{\imath}}) \ , \quad \forall l \right\} \quad i \in \mathcal{I} \end{aligned} \tag{8}$$

Similarly, the optimal codebook $\mathcal{C}^* = \{C_1^*, C_2^*, ..., C_M^*\}$ for a fixed partition is given by [7]

$$C_{\hat{\imath}}^* = \frac{\displaystyle\sum_{i=1}^{M} P(\hat{\boldsymbol{\imath}}|i) \int_{S_i} X p(X) \mathrm{d}X}{\displaystyle\sum_{i=1}^{M} P(\hat{\boldsymbol{\imath}}|i) \int_{S_i} p(X) \mathrm{d}X} \quad \hat{\boldsymbol{\imath}} \in \mathcal{I} \tag{9}$$

As it is shown in [8], $D(X, Y)$ is a finite sum of $d(\boldsymbol{x}, \boldsymbol{y})$, which is convex and differentiable, thus $D(X, Y)$ has the same properties. Therefore, the problem of minimizing the average distortion $\mathcal{D}(\mathcal{S}, \mathcal{C})$ is identical to the COVQ design problem but with a matrix distortion measure. A successive application of (8) and (9) results in a sequence of encoder-decoder pairs which converges to a local minimum as the LBG [3] and the COVQ algorithms do.

2.2. Optimization for a slow-fading Rayleigh Channel

Under the assumption that the channel is a slow-fading Rayleigh channel, optimum expressions (8) and (9) are still valid with the only difference that transition probabilities are, in this case, functions of the received SNR, ν, (the channel SNR, CSNR). Therefore, to compute the average distortion of the system we have to use average values of transition probabilities over all values of the the received SNR. In other words, we have to compute

$$\overline{P(j|i)} = \int_0^\infty P(j|i) p(\nu) \mathrm{d}\nu \quad i, j \in \mathcal{I} \tag{10}$$

where $P(j|i)$ are transition probabilities for an AWGN channel and $p(\nu)$ is the pdf of ν [6].

3. COMQ for LSP Parameters

Following the notation used in [9], LPC analysis is applied to speech frames of T ms of duration yielding LPC coefficient vectors $\boldsymbol{a}(m)$ which are transformed to LSP vectors $\boldsymbol{l}(m) = \left[l_1^m, l_2^m, \ldots, l_p^m\right]$ where p is the order of LPC filter. This process is performed over N consecutive speech frames to produce a $p \times N$ LSP matrix

$$X(m) = \begin{bmatrix} l_p^m & l_p^{m+1} & \ldots & l_p^{m+N-1} \\ \vdots & \vdots & \ldots & \vdots \\ l_2^m & l_2^{m+1} & \ldots & l_2^{m+N-1} \\ l_1^m & l_1^{m+1} & \ldots & l_1^{m+N-1} \end{bmatrix} \tag{11}$$

The matrix $X(m)$ is split into J submatrices with a general form given by

$$L_j(m) = \begin{bmatrix} l_{S(j)-1}^m & l_{S(j)-1}^{m+1} & \cdots & l_{S(j)-1}^{m+N-1} \\ \vdots & \vdots & \cdots & \vdots \\ l_{S(j-1)}^m & l_{S(j-1)}^{m+1} & \cdots & l_{S(j-1)}^{m+N-1} \end{bmatrix} \quad j = 1, \ldots, J \tag{12}$$

each with $r(j)$ rows. Thus, (11) can be expressed as

$$X(m) = \left[L_J(m), L_{J-1}(m), \ldots, L_1(m)\right]^T \tag{13}$$

In this work, each submatrix $L_j(m), \quad j = 1, \ldots, J$ is quantized using COMQ. We study the performance of implementing a Split COMQ for LSP quantization using a configuration used for SMQ in [9]. In particular, we have considered that $N = 4$ and $J = 10$ as the configuration to be studied. For comparison purposes, the performance of others quantization schemes are included. Specifically, we have included performance results of performing scalar quantization (SQ), 5-way Split VQ (SVQ), with 2 LSP parameters per subvector, and SMQ with $N = 4$ and $J = 10$ [9].

Table 1 shows studied quantization techniques and their characteristics. Speech frames are 30 ms long and a 10th order LPC filter is used. A weighted LSP distortion measure (IHM distance) [2] is used in the quantization process.

	bits/ frame	bits assignment	bits/ second
SQ	34	{3,4,4,4,4,3,3,3,3,3}	1333.3
SVQ	34	{7,8,7,6,6}	1333.3
SMQ	18.75	{9,9,8,8,8,8,8,7,6,4}	625
Split COMQ	18.75	{9,9,8,8,8,8,8,7,6,4}	625
$T = 30$ ms; $p = 10$			

Table 1: Analysis characteristics and number of bits per frame for the different coders.

4. Results and discussion

In this Section results on the performance of the considered LSP quantization techniques at different CSNRs are reported. Average spectral distortion (SD) is used as performance measure. Table 2 shows results for the average SD. In this table, row marked as SQ shows performance results when a scalar quantization is applied to the LSP parameters. Row marked as SVQ gives performance results for the 5-way SVQ. The SMQ experiment denotes a Split MQ of LSP parameters with the Generalized LBG algorithm. Rows marked as COMQ-21, COMQ-12, COMQ-6 and COMQ-0 show performance results for our technique, Split COMQ, in which the COMQ quantization codebooks are trained at a CSNR of 21, 12, 6 and 0 dB, respectively. Figure 2 shows average SD results graphically.

We have used 960 sentences from TIMIT database for training the quantization codebooks and 192 sentences out of training from TIMIT database to measure the performance of the simulated coders. For COMQ codebook design four CSNR (21, 12, 6 and 0 dB) have been considered. Performance results are obtained simulating the channel models at CSNR values of 21, 12, 6 and 0 dB and using GMSK modulation.

From Table 2 it can be observed that, in general, performance results of experiment SQ or SVQ are better than the performance of others experiments for slight to moderate noisy channels, but it has to be considered that SQ and SVQ use more bits per frame. However, SQ and SVQ suffer a bigger degradation as the channel noise increases. For example, the difference in average SD between CSNR of 21 dB and 0 dB is of 5.8 dB for SVQ while for COMQ-21 and COMQ-0 the difference is of 4.6 and 1.5 dB, respectively, when an AWGN Channel model is used. For slow-fading Rayleigh Channel, these differences are 6.5, 4.8 and 1.6 dB, for SVQ, COMQ-21 and COMQ-0 respectively. The percentage of outliers show the same behavior that the average SD shows.

Results show that an experiment COMQ-X gets the best performance at a CSNR at which design condition matches the channel condition. For example, at a CSNR of 0 dB, experiment COMQ-0 gives the best performance results compared to the others experiments, for both channel models. It is worth to note that COMQ-0 gives the best performance results at a bit-rate less than half the bit-rate of SQ or SVQ experiments. At a CSNR of 6 dB, experiment COMQ-6 gives the best performance results compared to the others experiments, in this case only for a slow-fading Rayleigh Channel model. For an AWGN Channel, COMQ-6 gives the best performance results compared to others COMQ-X or SMQ experiments.

The problem of a bigger complexity with MQ is increased with COMQ. But this drawback is mitigated due the presence of null cells in the quantization codebooks [1] when they are trained at certain noise level.

5. Summary

We have studied a joint source-channel coding technique, Split COMQ, applied to LSP parameters to reduce the bit rate and when the transmission is over a waveform channel. In this paper, the fundamentals of COMQ and its application to the coding of LSP parameters have been presented. Split COMQ has been used to remove intraframe/interframe redundancy of LSP parameters in such way that error protection is included without increasing the bit-rate.

Results show that for slight to moderate noisy channels SQ and SVQ give the lowest average SD but with a bit-rate greater than COMQ or SMQ. However, at certain noise levels in the channel, COMQ outperforms SQ and SVQ with the advantage of a much lower bit-rate transmission.

6. Acknowledgement

This paper has been supported by the Spanish CICYT project TIC-99-0583.

7. References

[1] N. Farvardin and V. A. Vaishampayan. On the Performance and Complexity of Channel-Optimized Quantizers. *IEEE Transactions on Information Theory*, 37; no 1:155–160, Jan. 1991.

[2] R. Laroia, N. P. Ayuso, and N. Farvardin. Robust and Efficient Quantization of Speech LSP Parameters Using Structured Vector Quantizers. In *ICASSP-91*, 1991.

[3] Y. Linde, A. Buzo, and R. M. Gray. An Algorithm for Vector Quantizer Design. *IEEE Transactions on Communications*, COM-28:84–95, Jan. 1980.

[4] K. K. Paliwal and B. S. Atal. Efficient Vector Quantization of LPC Parameters at 24 bit/frame. In B. S. Atal, V. Cuperman, and A. Gersho, editors, *Speech and Audio Coding for Wireless and Network Applications*, pages 191–201. Kluwer, 1993.

[5] J. L. Pérez-Córdoba, A. J. R. Ayuso, J. M. López-Soler, and V. E. Sánchez. Channel Optimized Matrix Quantization (COMQ) of LSP Parameters Over Waveform Channels. In *ICASSP-01 (Accepted)*, Salt Lake City, UT (USA), May 2001.

[6] J. G. Proakis. *Digital Communications*. McGraw-Hill, second edition, 1989.

[7] F. K. Soong and B. H. Juang. Line Spectrum Pairs (LSP) and Speech Data Compression. *Proceedings of ICASSP-84*, pages 1.10.1–1.10.4, 1984.

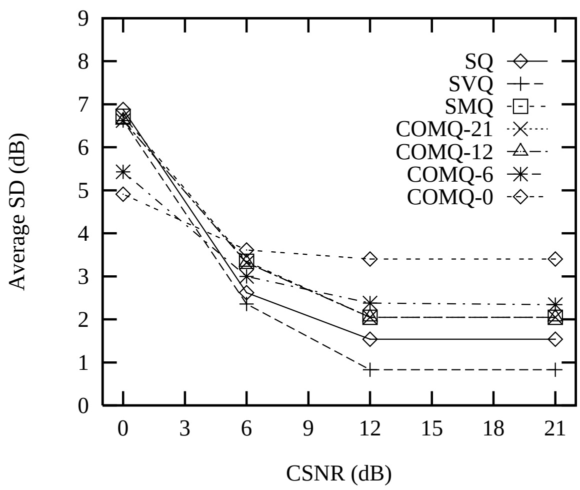
[8] C. Tsao and R. M. Gray. Matrix Quantizer Design for LPC Speech Using the Generalized Lloyd Algorithm. *IEEE Transactions on Acoustic, Speech and Signal Processing*, ASSP-33(3):537–545, June 1985.

[9] C. S. Xydeas and C. Papanastasiou. Split Matrix Quantization of LPC Parameters. *IEEE Transactions on Speech and Audio Processing*, 7(2):113–125, Mar. 1999.

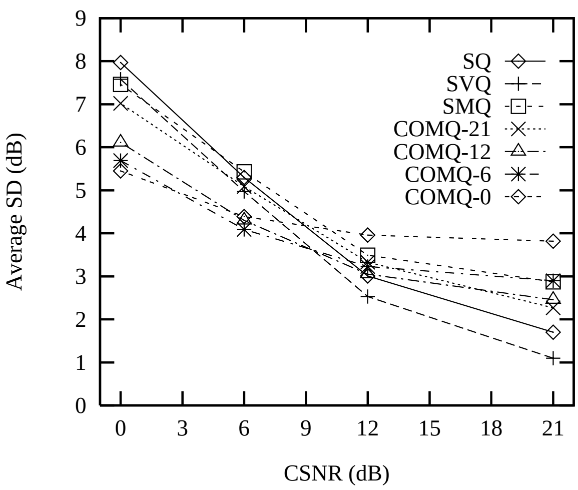
	CSNR (in dB)	Av. SD (in dB)	Outliers (in %) 2-4 dB	Outliers (in %) > 4 dB	CSNR (in dB)	Av. SD (in dB)	Outliers (in %) 2-4 dB	Outliers (in %) > 4 dB
SQ	21	1.54	12.66	0.23	21	1.70	16.69	2.83
	12	1.54	12.66	0.23	12	3.01	33.76	26.59
	6	2.62	30.97	19.97	6	5.30	24.93	66.39
	0	6.88	10.50	88.25	0	7.97	3.60	96.35
SVQ	21	0.83	0.37	0.00	21	1.10	4.99	3.00
	12	0.83	0.46	0.03	12	2.53	23.48	21.42
	6	2.36	21.90	18.95	6	4.97	26.16	60.19
	0	6.62	14.74	82.38	0	7.58	8.42	90.82
SMQ	21	2.05	45.07	1.33	21	2.87	49.23	5.65
	12	2.05	46.57	1.40	12	3.49	50.89	32.21
	6	3.35	52.06	28.75	6	5.43	22.51	75.92
	0	6.72	7.64	92.23	0	7.46	3.51	96.47
COMQ-21	21	2.05	45.14	1.29	21	2.27	49.29	5.15
	12	2.05	45.24	1.35	12	3.33	53.78	28.08
	6	3.32	52.43	28.31	6	5.09	27.45	70.55
	0	6.62	7.92	91.93	0	7.02	4.57	95.40
COMQ-12	21	2.05	45.25	1.35	21	2.46	61.28	5.70
	12	2.05	45.34	1.41	12	3.05	65.38	17.85
	6	3.32	52.48	28.35	6	4.30	43.98	53.11
	0	6.62	7.95	91.89	0	6.11	11.41	88.39
COMQ-6	21	2.34	60.53	4.12	21	2.90	66.43	13.87
	12	2.38	60.57	4.15	12	3.23	65.50	21.91
	6	3.00	65.78	16.77	6	4.09	49.68	46.65
	0	5.43	20.26	79.26	0	5.69	17.20	82.52
COMQ-0	21	3.40	63.36	26.38	21	3.82	57.09	37.61
	12	3.40	63.38	26.39	12	3.96	54.29	41.45
	6	3.61	60.66	32.07	6	4.39	43.39	54.41
	0	4.91	31.10	67.92	0	5.45	22.10	77.54

(a) **(b)**

Table 2: Average spectral distortion for different CSNR and different coders: (a) AWGN Channel, (b) Slow-fading Rayleigh Channel.



(a)



(b) Channel

Figure 2: Average SD: (a) AWGN Channel, (b) Slow-fading Rayleigh Channel.