CHANNEL OPTIMIZED MATRIX QUANTIZATION (COMQ) OF LSP PARAMETERS OVER WAVEFORM CHANNELS

José L. Pérez-Córdoba and Antonio Rubio and Juan M. López-Soler and Victoria E. Sánchez

Dpto. Electrónica y Tecnología de Computadores
Facultad de Ciencias. Universidad de Granada (Spain)
Phone: +34-958-243271 Fax: +34-958-243230 E-mail: jlpc/rubio/juanma/victoria@ugr.es

ABSTRACT

Combined source and channel coding is a technique to mitigate channel errors without increasing the bit error rate. Channel optimized vector quantizer (COVQ) [3] performs these objectives in the context of vector quantization. This paper presents a study of channel optimized matrix quantizer (COMQ) applied to quantize the Line Spectral Pair (LSP) parameters [5] as an extension of COVQ technique. Gaussian and slow-fading Rayleigh channels are considered and GMSK (Gaussian Minimum Shift-Keying) is used as modulation technique. Several channel signal to noise ratio (CSNR) are considered to measure the performance of this system. In addition, for comparison purposes, the performance of other schemes for quantizing the LSP parameters are computed.

Keywords: Matrix quantization, joint source-channel coding, COVQ, COMQ, LSP parameters, CELP coders, SMQ, GSM EFR coder.

1. INTRODUCTION

The LSP parameters are generally used to represent the short-time speech spectrum and are widely used in several international coder standards, as the DoD FS-1016 standard [1] or the GSM Enhanced Full Rate (EFR) coder [2]. Usually, the performance of these coders degrade in the presence of channel errors, therefore redundant information have to be added to protect the data against channel errors. There are several approaches in which a joint source and channel coding is performed. One of these approaches is COVQ which, as it was mentioned earlier, reduces redundancy in the source and protectes against errors at the same time.

In [6] a study of matrix quantization (MQ) for speech signal is reported, though without considering channel errors.

In the present work, we extend the COVQ technique to the matrix case, resulting in the Channel Optimized Matrix Quantization (COMQ) and this technique is applied to the coding of the LSP parameters.

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

2. COMQ TECHNIQUE

In this Section we present the fundamentals of COMQ technique and necessary optimal conditions are obtained. Although we are interested in speech signals, 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 kxN matrix M-level MQ and a waveform channel, an Additive White Gaussian Noise (AWGN) channel.

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 $\gamma : \mathbb{R}^{\mathrm{N}} \times \mathbb{R}^{\mathrm{k}} \rightarrow \mathcal{I}$, where $\mathcal{I} = \{1, 2, \ldots, M\}$, is described in terms of a partition $\mathcal{S} = \{S_1, S_2, ..., S_M\}$ of $\mathbb{R}^{\mathrm{N}} \times \mathbb{R}^{\mathrm{k}}$ 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. Specifically, we assume that we have an elementary signal constellation $\mathcal{T} = \{\boldsymbol{t}_1, \boldsymbol{t}_2, ..., \boldsymbol{t}_P\}$, consisting of P signals each one of dimension L_1, $\boldsymbol{t}_j \in \mathbb{R}^{\mathrm{L}_1}, \;\; j = 0, 1, ..., P$. Let us assume that $M = P^{L_2}$. Then, the signal $\boldsymbol{s}$ to be transmitted is selected from an expanded signal constellation $\mathcal{S} = \mathcal{T}^{L_2}$, the

This work was supported by the CICYT TIC99-0583

L_2-fold Cartesian product of $\mathcal{T}$. The effective dimension of the expanded signal constellation is $L = L_1 L_2$. The signal $\boldsymbol{s}(i)$ used for encoding the index i is given by

$$\boldsymbol{s}(i) = \left(\boldsymbol{t}^{(i_1)}, \boldsymbol{t}^{(i_2)}, ..., \boldsymbol{t}^{(i_{L_2})}\right) \tag{2}$$

where $(i_{L_2} i_{L_2-1} ... i_2 i_1)$ is the representation of i in base P. We restrict our study to BPSK modulation, so $L_1 = 1$ and $L_2 = L$.

The channel is a AWGN channel. 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

$$r_l = s_l + n_l, \qquad l = 1, 2, ..., L \tag{3}$$

where n_l's are i.i.d. zero-mean Gaussian random variables with common variance $\sigma^2 = N_0/2$. $N_0/2$ is the one-sided spectral density of the noise.

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 $\hat{\boldsymbol{\imath}}$, 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

$$\beta(\hat{\boldsymbol{\imath}}) = \beta\left(\hat{\boldsymbol{\imath}}(\boldsymbol{r})\right) = C_{\hat{\boldsymbol{\imath}}}, \quad C_{\hat{\boldsymbol{\imath}}} \in \mathbb{R}^{\mathrm{N}} \times \mathbb{R}^{\mathrm{k}} \quad \hat{\boldsymbol{\imath}} \in \mathcal{I} \tag{4}$$

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

$$\mathcal{D}(\mathcal{S}, \mathcal{C}) = \frac{1}{k} E\left[D\left(X, \beta\left(\hat{\boldsymbol{\imath}}(\boldsymbol{r})\right)\right)\right] \tag{5}$$

where $E[\cdot]$ means the expectation value and $D(X, Y)$ means the distortion measure used in the Generalized Linde-Buzo-Gray (GLBG) algorithm [6] defined by

$$D(X, Y) = \frac{1}{N} \sum_{n=1}^{N} d(\boldsymbol{x}_n, \boldsymbol{y}_n) \tag{6}$$

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

$$R = \frac{1}{kN} \log_2 M \quad \text{bits/sample} \tag{7}$$

The average distortion is a generalization to matrix quantization of the average distortion given in [5] for COVQ and it is given by

$$\mathcal{D}(\mathcal{S}, \mathcal{C}) = \frac{1}{k} \sum_{i=1}^{M} \int_{S_i} p(X) \left\{ \sum_{\hat{\boldsymbol{\imath}}=1}^{M} P(\hat{\boldsymbol{\imath}}|i) D(X, C_{\hat{\imath}}) \right\} \mathrm{d}X \tag{8}$$

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 [3] and from (8) it becomes clear that for a fixed $\mathcal{C}$, the optimum partition $\mathcal{S}^* = \{S_1^*, S_2^*, ..., S_M^*\}$ is given by

$$\begin{aligned} S_i^* \;=\; \Bigg\{ X \;:\; & \sum_{\hat{\boldsymbol{\imath}}=1}^{M} P(\hat{\boldsymbol{\imath}}|i) D(X, C_{\hat{\imath}}) \\ \leq\; & \sum_{\hat{\boldsymbol{\imath}}=1}^{M} P(\hat{\boldsymbol{\imath}}|l) D(X, C_{\hat{\imath}}) \;, \quad \forall l \Bigg\} \quad i \in \mathcal{I} \end{aligned} \tag{9}$$

Similarly, the optimal codebook $\mathcal{C}^* = \{C_1^*, C_2^*, ..., C_M^*\}$ for a fixed partition is given by [5]

$$C_{\hat{\imath}}^* = \frac{\sum_{i=1}^{M} P(\hat{\boldsymbol{\imath}}|i) \int_{S_i} X p(X) \mathrm{d}X}{\sum_{i=1}^{M} P(\hat{\boldsymbol{\imath}}|i) \int_{S_i} p(X) \mathrm{d}X} \quad \hat{\boldsymbol{\imath}} \in \mathcal{I} \tag{10}$$

As it is shown in [6] $D(X, Y)$ 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 (9) and (10) results in a sequence of encoder-decoder pairs which converges to a local minimum as the LBG ([4]) and the COVQ algorithms do.

2.2. COMQ for LSP Parameters

To obtain the LSP parameters we perform a LP analysis similar to the analysis performed in the GSM EFR standard coder. In the GSM EFR coder a LP analysis is performed twice per frame using two different asymmetric windows. Both sets of LP coefficients are quantified using the LSP representation. In that coder, a first order MA prediction is applied and the two residual LSP vectors are jointly quantized using split matrix quantization (SMQ) [7]. The matrix of the two residual vectors is split into 5 submatrices of dimension 2x2 (two elements from each vector).

In this work, we study the performance of implement a Split COMQ for LSP quantization. As in the EFR coder, we split the matrix of the two residual LSP vectors into 5 submatrices. These are quantified with COMQ with 7, 8, 7, 6 and 6 bits respectively, so that the number of bits per frame for the spectrum information is the same as in FS-1016 standard. A weighted LSP distortion measure is used in the quantization process. The weighting factors are calculated as in GSM EFR coder.

3. EXPERIMENTS

Three different kind of experiments are considered, shown in Table 1. CELP experiment carries out an independent scalar quantization of LSP parameters, in the same way it is done in CELP FS-1016 standard coder. GLBG experiment denotes a Split MQ of the residual LSP vectors with the Generalized LBG algorithm. This is done as in the application of COMQ technique to LSP quantization described above. Finally, COMQ-X experiment represents the application of COMQ technique to LSP quantization in which quantization codebooks are trained at a CSNR of X dB.

We have used 960 files from TIMIT database for training GLBG and COMQ quantization codebooks and 192 files 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.

4. RESULTS AND DISCUSSION

In this Section results on the performance of the considered LSP quantization techniques are reported. Average spectral distortion is used as performance measure. Table 2 shows results for the average spectral distortion (SD). In this table, row marked as CELP shows performance results when a scalar quantization is applied to the LSP parameters. Row marked as GLBG gives performance results for the split MQ. Rows marked as COMQ-21, COMQ-12, COMQ-6 and COMQ-0 show performance results for our technique in which COMQ quantization codebooks are trained at a CSNR of 21, 12, 6 and 0 dB, respectively.

From Table 2 it can be observed that in general performance results of experiment CELP are worse than performance of others experiment, and the performance difference grows with an increment of the noise in the channel. The exception is when the training of quantization codebooks is done under a very noisy channel condition. This fact occurs, for example, with COMQ-0 experiment considering a Gaussian Channel at a CSNR of 21 o 12 dB. Considering a slow-fading Rayleigh Channel, the mentioned situation appears in COMQ-6 and COMQ-0 at a CSNR of 21 dB. Regarding to the percentage of outliers the same conclusions can be established.

Results show that an experiment gets the best performance at a CSNR at which design condition matches the channel condition. For example, at a CSNR of 6 dB, experiment COMQ-6 gives the best performance results compared to the others experiments, for both channel models. From Table 2 it is clear that COMQ-X coders outperform other considered coders, specially for a noisy channel. For example, for a Gaussian Channel at a CSNR of 6 dB, COMQ-12 gives a 0.02 dB reduction in the average SD compared with GLBG, but at a CSNR of 0 dB this difference in performance is of 0.19 dB. Considering a slow-fading Rayleigh Channel, these differences in performance are now of 1.25 and 1.99 dB, respectively

However, COMQ-X experiments have a important drawback. This drawback is a bigger computational complexity. But this complexity is mitigated by the fact of presence of null cell in the quantization codebook [3] when a noisy channel is considered.

5. SUMMARY

We have studied a novel joint source-channel coding technique applied to LSP parameters when transmitting them over a waveform channel. At the same CSNR, the simulation shows that is possible to achieve a small cepstral distortion by mean of COMQ when a noisy channel is considered. The performance reported shows that COMQ technique can be a good technique for noisy channels as in wireless communications.

6. REFERENCES

[1] J. Campbell, T. Tremain, and V. Welch. An Expandable Error-Protected 4800 bps CELP Coder (U.S. Federal Standard 4800 bps Voice Coder). *Proceedings of ICASSP-89*, pages 735–738, 1989.

[2] E.T.S.I. International Standard GSM 06.60 (ETS 300 726). "Enhanced Full Rate (EFR) speech transcoding".

[3] 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.

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

[5] 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.

[6] 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.

[7] C. Xydeas and C. Papanastasiou. Efficient Coding of LSP Parameters Using Split Matrix Quantization. *Proceedings of ICASSP-95*, pages 740–743, 1995.

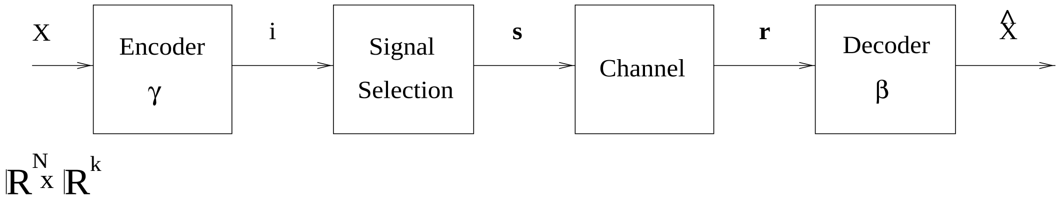

Figure 1: Block diagram of the COMQ system.

Coder	CELP	GLBG	COMQ-X
Update	30 ms	30 ms	30 ms
Order	10	10	10
Analysis	Open loop; Correlation; 15 Hz BW exp; Hamming window 30 ms; no preemphasis	Same as in CELP Coder	Same as in CELP Coder
Bits/frame	34, indep. LSP {3,4,4,4,4,3,3,3,3,3}	34, Split MQ of LSP {7,8,7,6,6}	34, Split COMQ of LSP {7,8,7,6,6}

Table 1: Characteristics of the spectral analysis and number of bits per frame for the different coders.

	(a) CSNR (in dB)	(a) Av. SD (in dB)	(a) Outliers (in %) 2-4 dB	(a) Outliers (in %) > 4 dB	(b) CSNR (in dB)	(b) Av. SD (in dB)	(b) Outliers (in %) 2-4 dB	(b) Outliers (in %) > 4 dB
CELP	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
GLBG	21	1.17	3.56	0.19	21	1.27	7.10	0.64
	12	1.17	3.56	0.19	12	2.61	38.17	16.20
	6	2.17	30.0	9.77	6	4.47	36.62	52.65
	0	6.15	18.24	80.76	0	7.18	8.33	91.45
COMQ-21	21	1.19	4.44	0.23	21	1.27	7.30	0.50
	12	1.19	4.44	0.23	12	2.42	38.41	11.99
	6	2.15	31.19	8.85	6	4.16	42.46	44.67
	0	5.97	19.98	78.94	0	6.60	12.54	87.16
COMQ-12	21	1.19	4.47	0.24	21	1.46	12.90	0.51
	12	1.19	4.47	0.24	12	2.00	36.94	2.59
	6	2.15	31.15	8.91	6	3.12	61.84	18.54
	0	5.96	20.15	78.73	0	5.19	30.60	68.14
COMQ-6	21	1.41	10.76	0.44	21	1.73	24.38	1.47
	12	1.41	10.76	0.44	12	2.07	40.57	2.95
	6	1.89	31.32	2.42	6	2.84	63.49	12.74
	0	4.51	43.62	53.19	0	4.57	41.91	55.65
COMQ-0	21	2.19	42.41	4.96	21	2.45	50.13	8.26
	12	2.19	42.41	4.96	12	2.60	55.73	9.78
	6	2.34	49.31	6.02	6	3.06	64.44	17.12
	0	3.78	57.90	35.42	0	4.34	46.92	49.91

(a) **(b)**

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