A STUDY OF JOINT SOURCE-CHANNEL CODING OF LSP PARAMETERS FOR WIDEBAND SPEECH CODING

José L. Pérez-Córdoba, Antonio M. Peinado, Victoria E. Sánchez, Antonio J. Rubio

Dpto. Electrónica y Tecn. de Computadores. Universidad de Granada. 18071 Granada (Spain)
Phone: +34-958-244011 Fax: +34-958-243230 E-mail: jlpc/amp/victoria/rubio@ugr.es

ABSTRACT

A study of combined source and channel coding applied to LSP parameters in wideband speech coding is presented. The traditional approach to protect against channel errors is to increase the bit-rate for channel coding, decreasing the bit-rate of the source coding according the channel conditions. Joint source-channel coding is an alternative that provides a technique to mitigate channel errors without an increase of the bit-rate due to channel coding. This paper presents a study of Channel Optimized Vector Quantizer and Channel Optimized Matrix Quantizer applied to Line Spectral Pairs (LSP) parameters in wideband speech coding. Gaussian and slow-fading Rayleigh channels are considered and GMSK (Gaussian Minimum Shift-Keying) is used as modulation technique. In addition, for comparison purposes, the performance of other schemes (Split Vector Quantization, Split Matrix Quantization and Split Multi-Stage Vector Quantization) for quantizing the LSP parameters are evaluated.

1. INTRODUCTION

In the last years, standards telecommunications bodies (3GPP/ETSI and ITU/T) have established a new standard wideband speech codec. The coder is referred as Adaptive Multi-Rate Wideband (AMR-WB) [1]. The wider bandwidth of this codec (7 kHz) gives improved speech quality and voice naturalness. AMR-WB is based on the ACELP (Algebraic Code Excitation Linear Prediction) algorithm and consists of nine speech codec modes with bit-rates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.6 kbps [1]. AMR-WB adapts the bit-rate allocation between speech and channel coding according the channel conditions. As the channel gets more noisy a lower bit-rate mode is used, dedicating more bits to channel coding. In order to avoid an increase of channel coding for error protection, joint source-channel coding techniques can be applied. One of these techniques is Channel Optimized Vector Quantization (COVQ) [2] in the context of vector quantization (VQ) and Channel Optimized Matrix Quantization (COMQ) [3] in the context of matrix quantization (MQ).

This work was supported by the CICYT TIC2001-3323

For speech coding, an efficient and a robust coding of the speech spectrum is needed. This information is usually represented using LSP parameters [4]. In narrowband speech coding, Split VQ (SVQ) or Split MQ (SMQ) is used to reduce complexity in the coding process of LSP parameters. In the wideband case, due to a larger vector size, Split Multi-Stage Vector Quantization (S-MSVQ)[1] is used.

All of the above quantization techniques don't consider channel errors so a degradation of speech signal occurs when noise appears. In this paper, a study of joint source-channel coding applied to the coding of LSP parameters in wideband speech coding is developed. Specifically, we study the performance of applying Split COMQ and Split COVQ to LSP parameters coding when transmission is over a waveform channel. For comparison purposes, we study the performance of SVQ, SMQ and S-MSVQ techniques.

This paper is organized as follows. In Section 2, an overview of the COMQ technique is given. COVQ technique will be presented as a special case of COMQ. Section 3 presents the application of COMQ and COVQ to LSP parameters coding. The performance results of the studied techniques are reported in Section 4. Finally, Section 5 contains a summary of this paper.

2. OVERVIEW OF COMQ AND COVQ TECHNIQUES

In this Section we give an overview of COMQ and COVQ techniques. To describe the COMQ technique we use the transmission system of Figure 1. We considere the source to be 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 Nxk matrix M-level MQ and a waveform channel.

As Figure 1 shows, the system consists of an 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 $\mathcal{I} =$

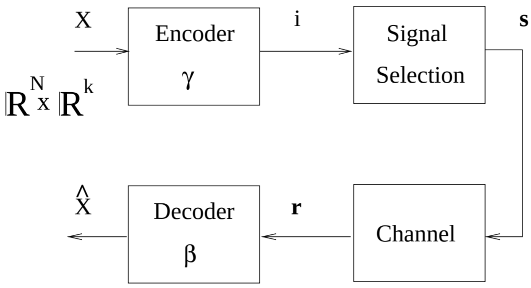


Fig. 1. Block diagram of the COMQ system.

$\{1, 2, \ldots, M\}$, 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 $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N)^T$ 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 into a signal $\boldsymbol{s}$ that is transmitted over the channel. The signal $\boldsymbol{s}$ is selected from a signal constellation of dimension L. The details of this block are given in [3].

We consider that the channel is an Additive White Gaussian Noise (AWGN) Channel. The random channel output vector $\boldsymbol{r}$ is related to the input vector $\boldsymbol{s}$ through

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

where 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) $\boldsymbol{r}$. Actually, the decoder β makes an estimate, $\hat{\boldsymbol{\imath}}$, of the index transmitted, i (hard-decision decoder). 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 describes the decoder through

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

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

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

where $E[\cdot]$ means the expected value and $D(X, Y)$ is given by $D(X, Y) = (1/N) \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{y}_n\|^2$. The encoding rate is given by $R = \frac{1}{kN} \log_2 M$ bits/sample.

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 $\mathcal{C}$. Through the minimization procedure, necessary conditions for optimality can be obtained [2][3]. These can be stated as follows. For a fixed $\mathcal{C}$, the optimum partition $\mathcal{S}^* = \{S_1^*, S_2^*, ..., S_M^*\}$ is given by

$$\begin{aligned} S_i^* \quad = \quad \Bigg\{ X \; : \; & \sum_{\hat{\boldsymbol{\imath}}=1}^{M} P(\hat{\boldsymbol{\imath}}|i) D(X, C_{\hat{\imath}}) \\ & \leq \sum_{\hat{\boldsymbol{\imath}}=1}^{M} P(\hat{\boldsymbol{\imath}}|l) D(X, C_{\hat{\imath}}) \,, \quad \forall l \Bigg\} \quad i \in \mathcal{I} \end{aligned} \tag{5}$$

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

$$C_{\hat{\imath}}^* = \frac{\sum_{i=1}^{M} P(\hat{\boldsymbol{\imath}}|i) \int_{S_i} X p(X) \mathrm{d}X}{\sum_{i=1}^{M} P(\hat{\boldsymbol{\imath}}|i) \int_{S_i} p(X) \mathrm{d}X} \qquad \hat{\boldsymbol{\imath}} \in \mathcal{I} \tag{6}$$

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.

The successive application of (5) and (6) results in a sequence of encoder-decoder pairs which converge to a local minimum as the LBG [5] algorithm does.

2.1. COVQ Technique

COVQ technique can be considered a special case of COMQ in which $N = 1$. The COVQ system will be described by an encoder $\gamma_V : \mathbb{R}^k \rightarrow \mathcal{I}$, and a decoder $\beta_V : \mathcal{I} \rightarrow \mathbb{R}^k$. Signal mapping and channel blocks have the same characteristics as the COMQ system. Optimum expressions for COVQ system are the equivalents to expressions (5) and (6) with $N = 1$.

2.2. Optimization for a slow-fading Rayleigh Channel

Under the assumption that the channel is a slow-fading Rayleigh channel, optimum expressions (5) and (6) are still valid with the only difference that transition probabilities are, in this case, functions of the received SNR, ν, which is a random variable in this channel model [6]. Therefore, to compute the average distortion of the system we have to use average values of transition probabilities over all values of the received SNR. In other words, we have to compute

$$\overline{P(j|i)} = \int_0^\infty P(j|i) p(\nu) \mathrm{d}\nu \quad i, j \in \mathcal{I} \tag{7}$$

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

Coder	SCOMQ/SMQ	SCOVQ/SVQ	S-MSVQ
Update	20 ms	20 ms	20 ms
Order	16	16	16
Analysis	Open loop; Autocorrelation; 60 Hz BW exp; Hamming window 30 ms;	Same as in SMQ Coder	Same as in SMQ Coder
Bits/frame	46, Split MQ of LSP {5,7,8,7,6,5,4,4}	46, Split VQ of LSP {5,7,8,7,6,5,4,4}	46, Split MSVQ of LSP {8,8,6,7,7,5,5}

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

3. COMQ AND COVQ FOR LSP PARAMETERS

In order to apply COMQ for LSP parameters coding, these are obtained by performing an LP analysis similar to the analysis performed in the GSM EFR standard coder [7]. As in the GSM EFR codec, an LP analysis is performed twice per frame using two different asymmetric windows. Both sets of LP coefficients are quantified using the LSP representation. A first order MA prediction is applied and the two residual LSP vectors are jointly quantized using split COMQ. The matrix of the two residual vectors is split into 8 submatrices of dimension 2x2 (two elements from each vector). These are quantified with COMQ with 5, 7, 8, 7, 6, 5, 4 and 4 bits respectively, so that the number of bits per frame for the spectrum information is the same as in the AMR-WB standard for the higher modes. A weighted LSP distortion measure is used in the quantization process. The weighting factors are calculated as in AMR-WB coder. For comparison purposes we have implemented a Split MQ [8] with the same LP analysis as we have described.

To apply COVQ for LSP parameters coding we perform the same LP analysis as in AMR-WB standard coder. The 16 dimension LSP vectors are quantized with an 8-way Split COVQ. The subvectors are quantified with 5, 7, 8, 7, 6, 5, 4 and 4 bits respectively, so that the number of bits per frame for the spectrum information is the same as in the AMR-WB standard for the higher modes. For comparison purposes we have implemented a Split VQ and a Split MSVQ [1] as in the AMR-WB codec.

We have used 960 speech files from the TIMIT database for training quantization codebooks and 192 files out of training also from the TIMIT database to measure the performance of the simulated coders. For COMQ and COVQ codebook design four Channel Signal to Noise Ratio (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. Five different experiments are considered as shown in Table 1.

Average spectral distortion (SD) is used as performance measure. Figure 2 and Table 2 show results for the average SD. The S-MSVQ experiment carries out a Split MSVQ of LSP parameters, in the same way as in the AMR-WB codec. SMQ denotes a Split MQ of the residual LSP vectors with the Generalized LBG algorithm. SVQ denotes a Split VQ of the residual LSP vectors. SCOVQ-X and SCOMQ-X experiments represent the application of Split COVQ and Split COMQ techniques, respectively, to LSP quantization when quantization codebooks are trained at a CSNR of X dB.

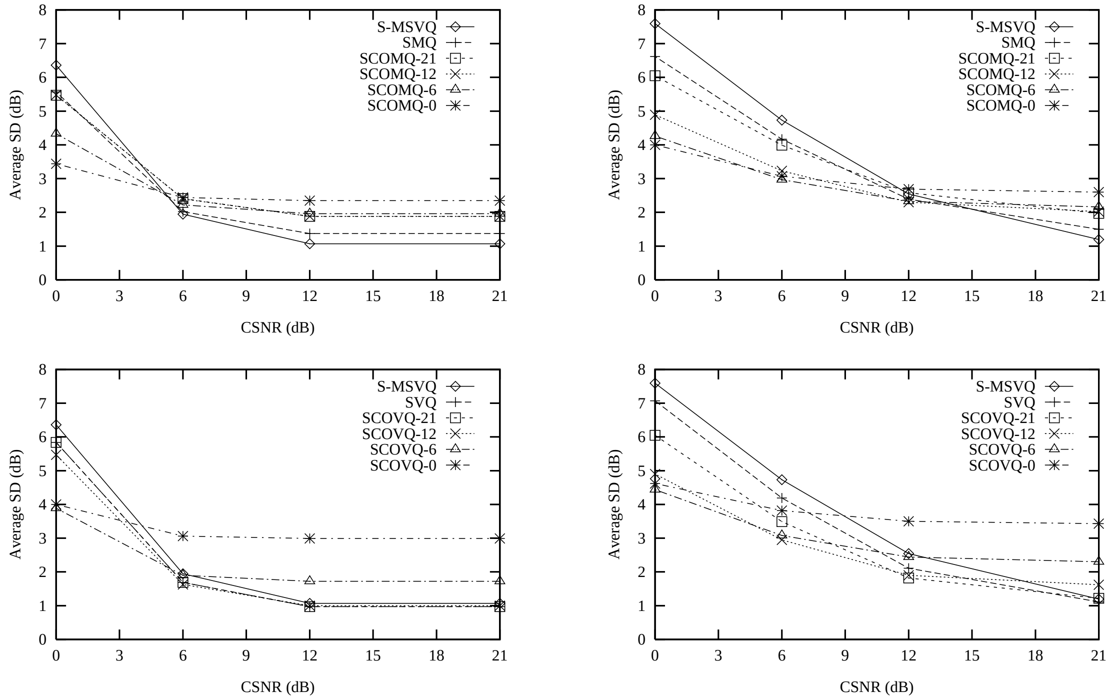
It can be observed in Figure 2 that joint source-channel coding techniques get better performance results under noisy channel conditions specially for a slow-fading Rayleigh Channel. For this channel model, under certain noise level, COVQ and COMQ techniques get better performance results compared to the rest of the studied techniques. It is worth while noticing that the performance of MSVQ degrades in a faster way than the other techniques when noise increases.

Comparing COMQ and COVQ techniques (Table 2), COMQ gets better performance results at lower CSNR and the differences in averaged SD are bigger for lower values of the CSNR.

Although COVQ and COMQ techniques show a bigger capacity to combat channel errors, they have two drawbacks. One of these is a channel mismatch condition which can be cancelled by choosing the most suitable coder/decoder pair to the channel conditions as in AMR-WB standard coder. The other one is a higher computational complexity but this is mitigated by the fact of the presence of null cells in the quantization codebook [2] when the channel is noisy.

5. SUMMARY

We have studied the performance results of several joint source-channel coding techniques applied to wideband speech LSP parameters coding when transmitting them over a waveform channel. Comparisons are made with Split VQ, Split MQ and Split MSVQ concluding that when channel



(a) (b)

Fig. 2. Average SD for different CSNR and different coders: (a) AWGN Channel; (b) Slow-fading Rayleigh Channel.

noise is present COVQ and COMQ get a smaller averaged spectral distortion. This is specially so when transmission is over a slow-fading Rayleigh Channel. COMQ is preferable over COVQ when the channel is noisier.

CSNR (dB)	21	12	6	0
S-MSVQ	1.20	2.54	4.73	7.60
SVQ	**1.11**	2.11	4.19	7.07
SMQ	1.50	2.39	4.17	6.62
SCOVQ-12	1.62	**1.90**	**2.95**	4.90
SCOMQ-12	2.02	2.30	3.23	**4.89**

CSNR (dB)	21	12	6	0
S-MSVQ	1.20	2.54	4.73	7.60
SVQ	**1.11**	**2.11**	4.19	7.07
SMQ	1.50	2.39	4.17	6.62
SCOVQ-6	2.30	2.44	3.08	4.44
SCOMQ-6	2.16	2.33	**2.97**	**4.26**

Table 2. Average SD for different CSNR. Comparison between some coders. Slow-fading Rayleigh Channel.

6. REFERENCES

[1] 3GPP, "Speech Codec speech processing functions; AMR Wideband Speech Code; Transcoding functions," 3GPP TS 26.190.

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

[3] J. L. Pérez-Córdoba, A. J. Rubio, J. M. López-Soler, and V. E. Sánchez, "Channel Optimized Matrix Quantization (COMQ) of LSP Parameters Over Waveform Channels," in *ICASSP-01*, May 2001.

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

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

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

[7] E.T.S.I., "Enhanced Full Rate (EFR) Speech Transcoding," International Standard GSM 06.60 (ETS 300 726).

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