INDEX ASSIGNMENT FOR PREDICTIVE WIDEBAND LSF QUANTIZATION

Vesa T. Ruoppila and Stéphane Ragot

University of Sherbrooke, Department of Electrical Engineering
Sherbrooke, Québec, J1K 2R1, Canada

Abstract **– In this paper we summarize some results derived earlier for the mean-square channel distortion of an *autoregressive moving average* (ARMA) vector quantizer with a maximum entropy encoder when the channel is assumed binary symmetric and memoryless. We discuss the required assumptions and their practical consequences in index assigment of ARMA vector quantizers. The discussion relates also to channel optimization of these quantizers. Furthermore, we compare noisy channel performance of memoryless, moving average, and autoregressive two-stage vector quantizers in line spectrum frequency quantization applied to wideband speech coding.**

1. INTRODUCTION

A predictor is customarily neglected when *index assignment* (IA) of an ARMA vector quantizer is considered. The channel distortion equation derived in [1] for a memoryless binary symmetric channel shows that this simplification may lead in some cases to a naively set IA problem. The derivation relies on the maximum entropy encoder assumption and requires that transmitted symbols at different time instants are statistically independent. However, empirical results in [1] indicate that the channel distortion expression is approximately valid in more practical conditions for small bit error probabilities. Therefore it provides a useful tool for assessing and optimizing the noisy channel performance of ARMA vector quantizers. Our objective is to extend this discussion by clarifying the assumptions needed and presenting new simulation results.

We will introduce notation and basic concepts in Section 2. In Section 3, we will then summarize briefly the analytical expression introduced in [1] for the channel distortion of ARMA vector quantizers. We will also show how this expression can be used in IA optimization and discuss the assumptions. These issues will be examined further in Section 4 in the context of *line spectrum frequency* (LSF) quantization applied to wideband speech coding. We will also compare the channel error resilience of several predictive two-stage vector quantizers. These simulations augment the viewpoint from the narrowband case in [1].

The reader is referred to [2] for an extensive discussion regarding predictive *vector quantization* (VQ), and [3] for a thorough tutorial on index assignment.

2. PRELIMINARIES

In this paper, we address the ARMA vector quantizer

$$\hat{\mathbf{y}}_S(t) = \sum_{k=1}^{n_A} \mathbf{A}_k \hat{\mathbf{y}}_S(t-k) + \sum_{k=0}^{n_B} \mathbf{B}_k \mathbf{u}_S(t-k) = \sum_{k=0}^{t} \mathbf{H}_k \mathbf{u}_S(t-k) \quad (1)$$

where $\hat{\mathbf{y}}_S(t)$ is the reconstruction of the source vector $\mathbf{y}(t)$ at the encoder and $\mathbf{u}_S(t)$ is the codevector at discrete time instant $t \geq 0$. The vectors $\hat{\mathbf{y}}_S(t)$, $\mathbf{y}(t)$ and $\mathbf{u}_S(t)$ are m-dimensional. The m by m coefficient matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ define the predictor. Usually $\mathbf{B}_0$ is set to an identity matrix. We utilize the impulse response representation for the predictor and denote the kth impulse response matrix by $\mathbf{H}_k$. The codevector $\mathbf{u}_S(t)$ is selected among the columns of the m by n codebook $\mathbf{U} = [\mathbf{u}_0\ \mathbf{u}_1\ \ldots\ \mathbf{u}_{n-1}]$.

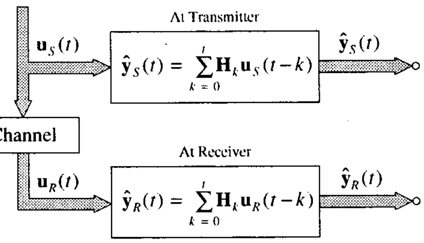

Figure 1. Subsystem of an ARMA vector quantizer.

The index of the selected codevector $\mathbf{u}_S(t) = \mathbf{u}_i$, $i \in I = \{0, 1, \ldots, n-1\}$, is transmitted over a channel to a decoder. The decoder receives the index $j \in I$ which may differ from i due to a channel error. The transmitted and received symbols are denoted by $s(t) = i$ and $r(t) = j$, respectively. The decoder reconstructs the quantized approximation $\hat{\mathbf{y}}_R(t)$ of the source using the received codevector $\mathbf{u}_R(t) = \mathbf{u}_j$. The system excluding the selection of $\mathbf{u}_S(t)$ is depicted in Fig. 1.

We describe the channel by its *transition probabilities* $p(j \mid i) = \mathrm{P}\{r(t) = j \mid s(t) = i\}$. They as well as the *transmission probabilities* $p(i) = \mathrm{P}\{s(t) = i\}$ are assumed time invariant. It should be pointed out that we assumed in (1) the source to be zero-mean for notational convenience. Moreover, we require (1) to be asymptotically stable. This requirement is always fulfilled in practice, since otherwise the system could not be realized.

3. CHANNEL DISTORTION IN ARMA VQ

The performance related to the transmission of information across the channel is measured by the *instantaneous channel distortion* $d(t) = \mathrm{E}\{\|\hat{\mathbf{y}}_S(t) - \hat{\mathbf{y}}_R(t)\|^2\}$ in terms of the squared Euclidean norm. This is the component of the instantaneous end-to-end distortion incurred only by channel errors. More particularly we are interested in the average performance over an infinite time interval,

$$D = \lim_{t \to \infty} (d(0) + d(1) + \ldots + d(t-1))/t. \quad (2)$$

The limit exists due to the posed stability and time-invariance assumptions. In [1], the channel distortion D was derived for maximum entropy encoders (for which $p(i) = 1/n$) assuming that the symbols transmitted at different time instants are independent. The main result of [1] is summarized in the following proposition.

The e-mail addresses of the authors are ruoppila@gel.usherb.ca and ragot@gel.usherb.ca.

This work was financed by VoiceAge Corp. and the NSERC.

0-7803-6416-3/00/$10.00©2000 IEEE

Proposition: Assume that a maximum entropy encoder and a memoryless binary symmetric channel produce symbols $s(t)$ and $r(t)$ that are independent of $s(t-k)$ and $r(t-k)$ for all $k \neq 0$. Then

$$D = \sum_{i,j \in I}\sum (\mathbf{u}_i - \mathbf{u}_j)^{\mathrm{T}} \mathbf{P} (\mathbf{u}_i - \mathbf{u}_j) p(i) p(j \mid i), \tag{3}$$

where

$$\mathbf{P} = \sum_{k=0}^{\infty} \mathbf{H}_k^{\mathrm{T}} \mathbf{H}_k. \tag{4}$$

The derivation of the proposition in [1] utilizes the equation

$$d(t) = \mathrm{tr} \sum_{k=0}^{t} \sum_{l=0}^{t} \mathbf{H}_k^{\mathrm{T}} \mathbf{H}_l \mathbf{C}_{k-l}, \tag{5}$$

in which $\mathbf{C}_k$ is the covariance matrix

$$\mathbf{C}_k = \mathrm{E}\{\mathbf{u}_\Delta(t) \mathbf{u}_\Delta^{\mathrm{T}}(t-k)\}, \quad k \in Z \tag{6}$$

and $\mathbf{u}_\Delta(t) = \mathbf{u}_S(t) - \mathbf{u}_R(t)$ is the error between the transmitted and the received codevector. In (5) tr stands for the trace of a matrix. A general expression of $\mathbf{C}_k$ needs a description for time dependencies remaining in the transmitted symbols sequence, but a more tractable expression is obtained utilizing the assumptions of the proposition. This also simplifies the channel distortion to the form (3).

The matrix **P** characterizes the effect of the predictor in (3). It can be factored with a square root decomposition, e.g., as $\mathbf{P} = \mathbf{P}^{1/2}\mathbf{P}^{1/2}$. Thus we get the transform $\mathbf{v}_i = \mathbf{P}^{1/2}\mathbf{u}_i$, and (3) becomes

$$D = \sum_{i,j \in I}\sum (\mathbf{v}_i - \mathbf{v}_j)^{\mathrm{T}} (\mathbf{v}_i - \mathbf{v}_j) p(i) p(j \mid i). \tag{7}$$

This is the well-known channel distortion equation of memoryless VQ, but for the rotated and scaled codebook $\mathbf{V} = \mathbf{P}^{1/2}\mathbf{U}$. Hence most results obtained for memoryless VQ can be generalized readily for (1). Obviously the index assignment design of memoryless VQ can be applied to the transformed codebook **V**.

However, we can often justifiably omit the predictor in IA design. Namely, (3) is bounded by

$$\Delta \le \lambda_{\min}\Delta \le D \le \lambda_{\max}\Delta, \tag{8}$$

in which $\lambda_{\min}$ and $\lambda_{\max}$ stand for the smallest and the largest eigenvalue of **P**. The symbol Δ denotes the channel distortion obtained by setting **P** in (3) to an identity matrix. The first inequality in (8) follows from the definition of **P**, which implies that $\lambda_{\min} \ge 1$. Equation (8) has an important implication; when all eigenvalues of **P** are equal, (3) returns to the classical channel distortion scaled by a scalar. Intuitively the gain of incorporating **P** in the IA design decreases as the ratio $\lambda_{\max}/\lambda_{\min}$ tends to one.

It should be noted that the discussion and results above are approximately valid also for non-maximum entropy encoders as the bit error probability is small in a sense its powers higher than one can be discarded. The property can be proved using this commonly employed assumption to the results of [1].

We also assumed the transmitted symbols $s(t)$ and $s(t-k)$ to be independent for $k \neq 0$. In other words, the transmitted symbol sequence is not allowed to contain residual information. This requirement is hard to meet fully in practice, albeit a properly designed predictor reduces significantly the time-dependency of a correlated source. The results in [1] demonstrate that (3) is able to characterize the empirical channel distortion at small bit error probabilities regardless the posed assumptions are not valid. Thus (3) provides a widely applicable expression for analyzing the channel distortion of ARMA VQ. The result can also used as a basis for deriving computational methods for improving the noisy channel performance of such systems.

Before proceeding to numerical examples, it should be observed that for scalar quantizers the proposition above returns to a classical result introduced in [4].

4. Numerical Examples

Experimental Setup. Linear prediction analysis of order 16 is done at every 20-ms frame using a Hamming window of 40 ms centered on the current frame. The input signal sampled at 16 kHz is band-pass filtered to 25–7000 Hz, and emphasized with the filter $H(z^{-1}) = 1 - 0.75z^{-1}$ prior to further processing. Autocorrelation coefficients are windowed to gain a bandwidth expansion of 60 Hz and white noise correction of 1e–4. After that the filter coefficients are solved with the Levinson-Durbin algorithm, and converted into the LSF representation for quantization.

Line spectrum frequencies are coded using a two-stage quantizer of 44 bits. The first stage is divided into nine- and seven-dimensional splits having eight and six bits, respectively. The second stage comprises five splits of six bits; the first four splits are three-dimensional and the last one is four-dimensional. Thus in total 14 bits are allocated for the first stage and 30 bits for the second one. All quantizers to be studied have the same split structure and bit allocation.

Codebooks and diagonal coefficient matrices for (1) are optimized sequentially applying the generalized Lloyd algorithm according to [5]. The estimation data comprises 149 000 frames corresponding to about 50 minutes of audio material. A separate data set of 25 000 frames is used for evaluating the performance. The databases comprise mainly speech in several languages, but also some music samples from different genres. We do not use a weighted norm in encoding, although it could be employed to improve the performance. The M-L search [5] is used in estimation with 16 survivors, and with 4 survivors for validation in order to reduce the complexity to a tolerable level.

Note that the discussion in Section 3 can be extended straightforwardly for multi-stage vector quantizers, albeit some further assumptions are needed. That is, all symbols related to different stages and their splits have to be independent. After this the splits can be considered separately.

We optimize index assignments of quantizers using the binary switching algorithm [6] with estimated transmission probabilities. The quality of a local minimum found by the algorithm affects considerably the performance. Hence we run the algorithm 2000 times starting from a random initial value, and select the best IA for use.

Spectral Distortion Statistics. Noiseless channel performance of some ARMA(n_A, n_B) vector quantizers is summarized in Table 1 in terms of the average spectral distortion SD. The percentage of frames in which spectral distortion exceeds 2 dB is

Table 1. Performance in a noiseless channel.

	SD [dB]	SD_2 [%]	$\lambda_{\min}$	$\lambda_{\max}$
MA(0)	1.08	1.89	1	1
MA(1)	0.91	0.78	1.07	1.38
MA(2)	0.92	0.72	1.07	1.46
AR(1)	0.86	1.14	2.43	4.41
ARMA(1,1)	0.85	0.93	1.87	3.53

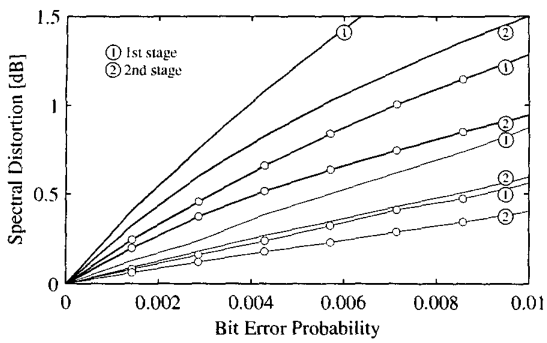

Figure 2. Empirical channel distortion for the MA(0) and ARMA(1, 1) quantizers shown by thick and thin lines, respectively, for both random and optimized IAs. Optimized IAs have been marked by circles.

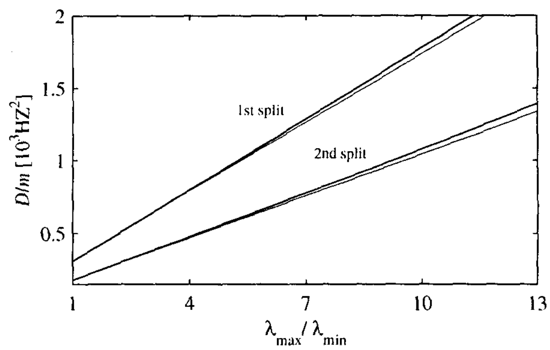

Figure 3. Normalized channel distortion (3) at the bit error probability 1e−3 for the first stage with (thin line) and without (thick line) the predictor in IA optimization.

denoted by SD_2. Spectral distortion, see e.g. [5] for the definition, is evaluated over the 50–7000 Hz band. Unstable filters are stabilized before computing the statistics by reordering the LSF parameters.

Noisy Channel Performance. The eigenvalues in Table 1 indicate that the ARMA(1, 1) quantizer may be very sensitive to channel errors. Therefore we select it for a more detailed case study. The noisy channel performance of this quantizers is compared in Fig. 2 to the memoryless vector quantizer. The channel is simulated using a uniform bit error probability e for all bits of one stage and keeping another stage error free. The results are averaged over 5 runs. Fig. 2 shows that IA optimization yields a significant performance gain over random indexing. For example, SD drops from 0.4 to 0.2 dB, and SD_2 from 6.6 to 3.1 % for the first stage at $e = 1.5\text{e}{-4}$. However, the quantizer is still considerably more sensitive to channel errors than the memoryless and MA quantizers listed in Table 1. They give the performance measures SD ≈ 0.1 dB and $SD_2 \approx 2$ %. The MA quantizers provide the best end-to-end performance until $e \approx 5\text{e}{-3}$, and lose the advantage to the memoryless quantizer after that.

We observed that the matrix **P** had a minor effect in the IA design. The performance loss caused by ignoring it was only a few percents at maximum. This is explained by relatively small values of $\lambda_{max}/\lambda_{min}$. Moreover, the initial index assignment used in the optimization affects considerably the outcome so direct comparisons are difficult. The performance gap between the best and worst minimum found in 2000 runs was more than 15 % in terms of the mean-square distortion used as the optimization criterion in the binary switching algorithm.

To assess relation between the channel distortion and the ratio $\lambda_{max}/\lambda_{min}$, we evaluate (3) separately for both splits in the first stage of the ARMA(1, 1) quantizer using a scaled **P** matrix. The codebooks are kept untouched. The kth eigenvalue is scaled as $\lambda_k := \lambda_{min} + r(\lambda_k - \lambda_{min})$ in which $r = (\lambda - \lambda_{min})/(\lambda_{max} - \lambda_{min})$ and λ is the greatest eigenvalue after scaling. This yields the desired value to the eigenvalue ratio but maintains λ_{min} constant. Other eigenvalues change linearly. Originally the series (4) gives the diagonal matrix $\mathbf{P} \approx \text{diag}\{1.9, 2.2, 2.7, 3.5, 2.9, 2.8, 2.7, 2.7, 3.0\}$ for the first split. The series (4) converges quickly for this predictor, since all poles are strongly inside the unit circle. It should be noted that this example is artificial and does not relate directly to the actual quantization problem.

Fig. 3 presents the normalized channel distortion D/m as a function of $\lambda_{max}/\lambda_{min}$ with and without incorporating the predictor in the IA design. The curves have been averaged over 200 runs. The performance gain is small but increases consistently with the eigenvalue ratio. However, the gain was over 10 % in some individual runs started from the same initial IA.

5. Conclusions

In this paper, we studied the error resilience of ARMA VQ and extended earlier discussion regarding the channel distortion. The results show that the predictor can often be omitted from the IA design, but this has to be verified by examining the eigenvalues of the **P**-matrix. Though this approximation is not always valid, its influence to the performance is small compared, for example, to the quality of a local minimum attained in IA optimization.

The channel distortion expression can be applied further to channel optimized VQ. Equation (3) is particularly useful for channel optimization of MA vector quantizers, since in this case channel distortion and coefficient matrices have a simple relation. In [7], channel optimization has been studied in the context of AR VQ. However, an optimization criterion has been derived by taking into account time dependency in transmitted symbols and simplifying the outcome using different assumptions than in this paper. The relation of these approaches deserves further examination.

References

[1] M. Vuolahti and V.T. Ruoppila, "Performance of predictive vector quantizers on noisy channels," *IEEE Nordic Signal Processing Symposium* (Norsig), Vigso, Denmark, pp. 213–216, 8–11 June 1998.

[2] T. Eriksson, J. Lindén, and J. Skoglund, "Interframe LSF quantization for noisy channels," *IEEE Trans. Speech and Audio Processing*, Vol. 7, No. 5, pp. 495–509, 1999.

[3] P. Hedelin, P. Knagenhjelm, and M. Skoglund, "Theory for transmission of vector quantization data," in *Speech Coding and Synthesis*, W.B. Kleijn and K.K. Paliwal, Eds., Elsevier Science, pp. 347–396, 1995.

[4] K.-Y. Chang and R.W. Donaldson, "Analysis, optimization, and sensitivity study of differential PCM systems operating on noisy communication channels," *IEEE Trans. Communications*, Vol. 20, No. 3, pp. 338–350, 1972.

[5] 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 Processing*, Vol. 1, No. 4, pp. 373–385, 1993.

[6] K. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Trans. Communications*, Vol. 38, No. 12, pp. 2147–2158, 1990.

[7] J. Lindén, *Interframe Quantization for Noisy Channels*, Doctoral Dissertation, Chalmers Univ. of Tech., Göteborg, Sweden, 1998.