PREDICTIVE VQ FOR NOISY CHANNEL SPECTRUM CODING: AR OR MA?

Jan Skoglund and Jan Lindén

Department of Information Theory
Chalmers University of Technology
S-412 96 Göteborg, Sweden.
jans@it.chalmers.se, linden@it.chalmers.se

ABSTRACT

In this paper, the performance of different predictive vector quantization (PVQ) structures is studied and compared for different degrees of channel noise. Predictive quantization schemes with an auto-regressive (AR) decoder structure are compared with schemes that employ a moving average (MA) decoder. For noisy channels MA prediction performs better than AR. It is shown here that a combination of a PVQ scheme (AR or MA) and a memoryless VQ outperforms both types of traditional predictive quantizer schemes in noiseless as well as noisy channels.

1. INTRODUCTION

In the strive for more efficient speech coding algorithms, there has been a growing interest in schemes exploiting interframe correlation in vector quantization (VQ) of, for example, the spectrum parameters [1-6]. An interesting aspect is how these methods perform in transmission over a noisy channel. Differential quantization or predictive quantization with an auto-regressive (AR) decoder structure has shown good performance for error-free transmission. However, the performance rapidly deteriorates when channel noise is introduced due to error propagation. Therefore, coding methods have been developed, where a trade-off in performance or complexity between abilities to cope with noise has been introduced by incorporating a moving average (MA) decoder. The error propagation is then limited by the non-recursive decoder structure.

In this work we compare the performance of these two different structures of predictive quantization paradigms. We also compare them with a combination structure, referred to as the safety-net method, which has shown promising results for noisy channels [5].

2. PREDICTIVE VECTOR QUANTIZATION

A predictive vector quantization scheme (PVQ[1]) exploits the memory of the input process by forming predictions of incoming vectors and then quantizing the prediction error, c.f. Figures 1 and 2. The prediction is based on previously quantized vectors.

[1] In the literature, the notation PVQ often stands for the case when the AR type of prediction is used. We will refer PVQ to the general case, incorporating both AR and MA predictors.

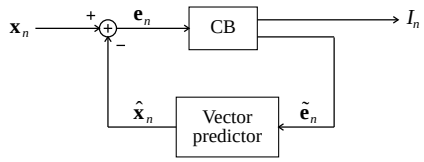

Figure 1. A PVQ encoder.

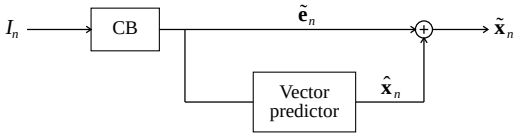

Figure 2. A PVQ decoder.

We will in this paper focus on two methods for linear prediction, the moving average and the auto-regressive approaches. In the MA case, the linear predictor, $\hat{\mathbf{x}}_n$, is given by

$$\hat{\mathbf{x}}_n = \sum_{k=1}^{M} \mathbf{B}_k \tilde{\mathbf{e}}_{n-k},$$

where $\tilde{\mathbf{e}}_{n-k}$ are previously coded prediction errors, $\mathbf{B}_k$ are prediction coefficient matrices and M is the predictor order.

An AR linear predictor can be written as

$$\hat{\mathbf{x}}_n = \sum_{k=1}^{M} \mathbf{A}_k \left(\tilde{\mathbf{e}}_{n-k} + \hat{\mathbf{x}}_{n-k} \right) = \sum_{k=1}^{M} \mathbf{A}_k \tilde{\mathbf{x}}_{n-k},$$

where $\tilde{\mathbf{x}}_{n-k}$ are previously coded vectors.

The final quantized vector, in both cases, is then the sum

$$\tilde{\mathbf{x}}_n = \hat{\mathbf{x}}_n + \tilde{\mathbf{e}}_n.$$

A general treatment of the concept of AR PVQ can be found in [7]. PVQ has previously been used with some success in spectrum coding. AR examples are [1, 5, 6] and examples with MA predictors are [2, 3].

In an error-free transmission situation, AR schemes have shown to perform better than MA schemes [2, 4]. The general argument for utilizing MA prediction is its advantages in transmission over a noisy channel. In an AR system, where the decoder has a recursive structure, a transmission error will not only affect the current decoded vector, but also (infinitely many) successive vectors. This error propagation will in the MA case be restricted to a few following vectors, determined by the predictor order.

3. SAFETY-NET VQ

The authors have previously described the concept of a *safety-net* extension to a memory VQ in [5, 8]. A memory VQ, such as PVQ, is combined with a fixed memoryless VQ, called the safety-net VQ, c.f. Figure 3.

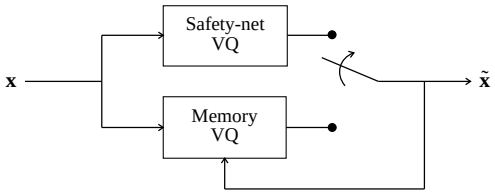

Figure 3. The safety-net principle: Combine a memory VQ with a fixed memoryless VQ (the safety-net VQ).

The search process is performed by first searching the memory VQ codebook for a candidate vector, then searching the fixed VQ codebook for a second candidate vector. The best candidate is then encoded and transmitted to the decoder. One of the advantages of a safety-net extended PVQ is better robustness against "outliers", i.e. vectors having low intervector correlation. In this scheme, the safety-net takes care of the outliers, and the PVQ can concentrate on stable segments with high interframe correlation.

Another advantage is obvious when data must be transmitted over a noisy channel: The error propagation that is evident in memory VQ schemes is significantly reduced, due to the frequent use of vectors from the memoryless safety-net quantizer. In this paper, we will examine safety-net extensions to both AR and MA PVQ schemes.

4. SIMULATIONS

The training database used in this work consists of 86 minutes of speech. Another database with a length of 7 minutes was used for evaluation. The speech was recorded from FM radio, low-pass-filtered at 3.4 kHz and sampled at 8 kHz. A 10th order LPC analysis using the stabilized covariance method with high-frequency compensation and error weighting (following [9]) was performed for each frame of speech. The prediction coefficients obtained from the analysis were then transformed to LSF parameters prior to quantization. A fixed 10 Hz bandwidth expansion was applied to each pole of the LPC polynomial. In this work, only the diagonal elements of the prediction matrices are used. This restriction has been employed in previous work on MA predictors [2-4]. We have experimentally found that the restriction only has a minor impact on performance for an AR system.

We have investigated two ways of determining the prediction coefficients for the MA predictors, a standard scalar method based on a high order AR approximation to the MA process [10] and the LMS type of algorithm presented in [3], without finding any measurable difference. The AR predictor coefficients are obtained by determining the full prediction matrices using the vector counterpart of the Levinson-Durbin algorithm (see e.g. [7]) and then using only the diagonal elements of the matrices.

We employ the weighted Euclidean distortion measure presented in [9] for the codebook search in all VQ techniques. For measuring the quantization performance, we calculate the spectral distortion (SD) in the 0-3 kHz band.

In this work we have utilized a 3-split VQ scheme for all quantizers, where the dimensions of the split vectors are 3, 3 and 4 respectively.

4.1. PREDICTOR ORDER

When comparing MA and AR prediction, one important difference is the predictor order required to obtain acceptable performance. In Figure 4, 24 bit AR and MA PVQs are compared for frame lengths of 10 ms and 20 ms. The bit allocation used for the three split vectors was (8,9,7). The length of the analysis window is 25 ms for both cases. For the AR predictors, using predictor order higher than one gives no noticeable performance gain while for the MA predictors predictor orders of say 3-5 are required to obtain acceptable performance. Note also that significantly better results are obtained if 10 ms frames are used (instead of 20 ms), due to the higher interframe correlation that results. Consequently, higher order predictors are needed for MA PVQ to reach the same performance as for AR PVQ.

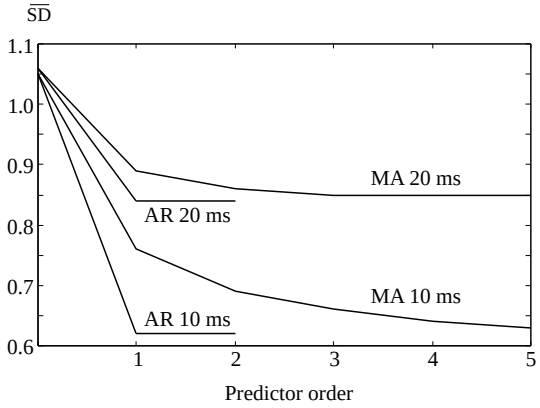

Figure 4. Performance comparison for different predictor orders and frame size (10 ms and 20 ms) at 24 bits. The analysis window is 25 ms.

Using a 25 ms analysis window for a frame size of 10 ms which results in a high interframe correlation but is not always realistic because of the extra delay imposed. Therefore, we have in Figure 5 compared PVQs for 10 ms frames with analysis windows of 12.5 ms and 25 ms. From this figure it is clear that the length of the analysis window has a significant impact on the interframe correlation and thereby the performance of PVQ schemes. Note that we here have used symmetrical windows, while asymmetrical windowing can obtain similar performance with less delay [2].

The simulation results reported in this section, as well as other investigations (e.g. [3]), indicate that higher order MA is needed to reach AR predictor performance. We have also demonstrated that PVQ performance improves with decreasing frame size, and also with increasing size of the analysis window. In the following experiments, we use

20 ms frames and 25 ms analysis windows, which is common in contemporary speech coders. From Figure 4 we observe that for this choice of analysis conditions, it suffices to employ third order MA and first order AR.

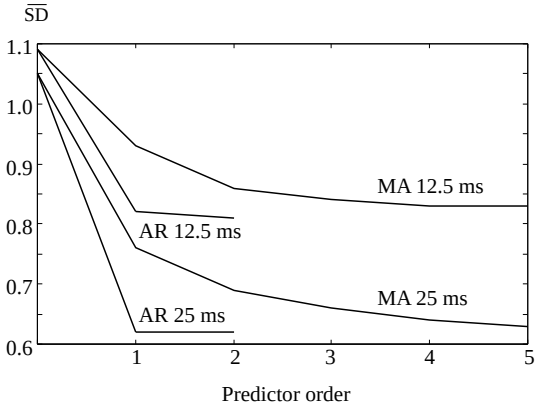

Figure 5. Performance comparison for different predictor order and analysis window size (12.5 ms and 25 ms) at 24 bits. The frame size is 10 ms.

4.2. NOISE-FREE CHANNEL PERFORMANCE

We have examined performance in terms of average spectral distortion for the following five quantizer schemes: Memoryless VQ (ML), AR PVQ with first order prediction (AR1), MA PVQ with third order prediction (MA3), AR PVQ with first order prediction and safety-net (SN-AR1) and MA PVQ with third order prediction and safety-net (SN-MA3). In Figure 6, the average SD of the investigated coding schemes is plotted as a function of the number of bits used. The figure verifies that PVQ methods can utilize interframe correlation and achieve performance significantly better than what is obtainable with memoryless VQ. It can also be seen that the schemes with AR predictors perform slightly better than the MA predictor schemes and that the safety-net extension yields an improvement of approximately 0.5 bits for MA PVQ and 1 bit for AR PVQ.

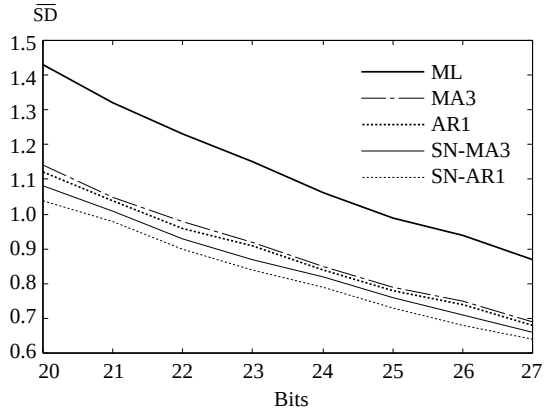

Figure 6. Performance of the VQ schemes as a function of codebook size.

In Table 1 the performance, both in average SD as well as outlier percentage, is presented for the five investigated coding methods at 24 bits. For the safety-net configurations, one bit was designated for indicating the chosen codebook, and the remaining 23 bits were allocated as (8,8,7). As also was found in [5], the introduction of a safety-net VQ does not only decrease the average distortion but also the number of outliers.

TABLE 1. QUANTIZER PERFORMANCE AT 24 BITS WITHOUT CHANNEL NOISE.

Quantizer	$\overline{SD}$ [dB]	2-4 dB [%]	>4 dB [%]
ML	1.06	0.90	0
MA3	0.85	0.90	0.005
AR1	0.84	0.87	0.015
SN-MA3	0.82	0.29	0
SN-AR1	0.79	0.25	0

4.3. NOISY CHANNEL PERFORMANCE

As previously stated, the main argument for utilizing MA prediction is its advantages over AR methods under noisy conditions. We will in this section verify this and also compare with the safety-net extended VQ methods.

We assume a memoryless binary symmetric channel with bit error probability q. Noisy channel performance of vector quantizers having random index assignment is in general poor. In this study, procedures to improve the index assignment are applied to all vector quantizers. We have applied a fast and reliable method described in [11].

The noisy channel performance for the VQs can be improved by "weakening" the prediction matrices as described in [5]. By decreasing the values of the prediction matrix elements, the performance for noise-free channels becomes slightly worse, but the performance for noisy channels is significantly improved. A simple improvement of the methods is to scale the prediction matrices with a constant $\mu<1$. We have experimentally found values of μ for each method resulting in an increase in the average SD of only 0.01 dB for noise-free conditions while giving large gains for high error rates.

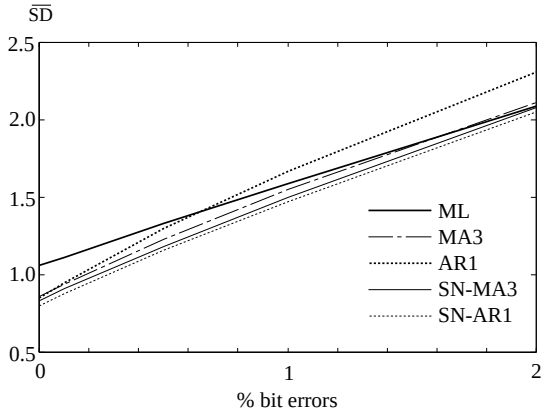

Figure 7. Performance of the VQs at 24 bits as a function of the channel bit error rate.

The average SD performance for all methods under equal channel conditions at 24 bits is depicted in Figure 7. In Table 2, the values of the simulations for q=1% are presented. As expected, the AR1 performance degrades at high error rates. We also note that the gains obtained by the predictive methods compared to memoryless VQ decreases at higher

noise levels. However, from these results we conclude that the safety-net extended methods are most robust to channel noise, with SN-AR1 performing better than SN-MA3.

TABLE 2. QUANTIZER PERFORMANCE AT 24 BITS AND 1% BIT ERROR RATE.

Quantizer	$\overline{SD}$ [dB]	2-4 dB [%]	>4 dB [%]
ML	1.59	12	6.7
MA3	1.55	18	4.7
AR1	1.67	21	5.4
SN-MA3	1.50	14	4.7
SN-AR1	1.47	15	4.6

To investigate what gains can be achieved in terms of bits employed for the same performance, we have in Table 3 examined all schemes for different number of bits. These results can be summarized as follows: Compared to memoryless VQ, 4 bits can be gained by using a safety-net extended AR PVQ, 3 bits employing a third order MA PVQ and 3.5 bits if the MA PVQ is extended with a safety-net. Note also that the AR PVQ without safety-net rapidly loses its advantage over memoryless VQ when the bit error probability is increased.

TABLE 3. $\overline{SD}$ COMPARISON FOR DIFFERENT BIT ERROR RATES.

Bit error [%]	20 bit SN-AR1	20 bit SN-MA3	21 bit AR1	21 bit MA3	24 bit ML
0	1.05	1.09	1.05	1.07	1.06
0.1	1.12	1.16	1.14	1.14	1.11
0.5	1.38	1.40	1.43	1.39	1.33
1	1.67	1.68	1.77	1.68	1.59
2	2.20	2.23	2.34	2.19	2.09
5	3.41	3.41	3.61	3.37	3.27

4.4. LISTENING TESTS

A subjective performance evaluation was obtained by conducting listening tests of preference. 21 test persons listened to 6 short sentences from the TIMIT database (3 male and 3 female speakers), and made pair-wise comparisons of the speech quality using headphones. In the test, the five quantizers in Table 3 was compared for the cases of 0% and 1% bit errors. Note that the quantizers have different sizes. Synthetic speech was generated by exciting the quantized production filter with the prediction error signal from the unquantized inverse filter. The LSF parameters were transformed into reflection coefficients and then linearly interpolated on a sample-by-sample basis.

A statistical evaluation, in the form of a series of t-tests [12], performed on the results revealed that there was a small general preference of the 20-bit SN-MA quantizer, and that the other four had comparable quality. For the noiseless case and a significance level of 5% SN-MA could be distinguished as better than the other memory VQs (not the memoryless VQ). For the noisy conditions the difference was more evident. The SN-MA quantizer could now be distinguished from all the others at a 1% significance level.

This means that although the five methods have comparable objective performance (Table 3), it may seem that the distortion from the SN-MA quantizer is the least annoying for a listener.

5. SUMMARY

We have in this paper found that even though MA PVQ outperforms AR PVQ for noisy channels, better performance can still be obtained. By extending a predictive VQ, either AR or MA, with a safety-net, the performance improves over the standard cases for all investigated bit error rates. The bit savings of the two safety-net extended predictive VQs are 1 bit compared to the standard AR and MA PVQs and 4 bits to a memoryless VQ. The performance difference between the safety-net extended AR and MA PVQs is small, SN-AR showed higher objective performance and the SN-MA was found to yield a higher subjective quality in the performed listening tests.

Hence, the safety-net extension is an advantageous alternative to traditional MA predictors for overcoming the problems connected to AR predictors in transmission over noisy channels.

6. REFERENCES

[1] S. Wang, E. Paksoy, and A. Gersho, "Product code vector quantization of LPC parameters," in *Speech and audio coding for wireless and network applications*, B. Atal, V. Cuperman, and A. Gersho, Eds., Kluwer Academic Publishers, pp. 251-258, 1993.

[2] R. Salami, C. Laflamme, J.-P. Adoul, and D. Massaloux, "A toll quality 8 kb/s speech codec for the personal communications system (PCS)," *IEEE Trans. Vehic. Technol.*, vol. 43, no. 3, pp. 808-816, 1994.

[3] H. Ohmuro, T. Moriya, K. Mano, and S. Miki, "Vector quantization of LSP parameters using moving average interframe prediction," *Electronics and Communications in Japan, Part 3*, vol. 77, pp. 12-26, 1994.

[4] L. Yin, "Performance comparison of LSP coding algorithms based on predictive vector quantization," *in Proc. ICSPAT '95*, Boston, USA, 1995, pp. 1878-1882.

[5] T. Eriksson, J. Lindén, and J. Skoglund, "Exploiting interframe correlation in spectral quantization - a study of different memory VQ schemes," *in Proc. ICASSP '96*, Atlanta, USA, vol. 2, 1996, pp. 765-768.

[6] H. Zarrinkoub and P. Mermelstein, "Switched prediction and quantization of LSP frequencies," *in Proc. ICASSP '96*, Atlanta, USA, vol. 2, 1996, pp. 757-760.

[7] V. Cuperman and A. Gersho, "Vector predictive coding of speech at 16 kbits/s," *IEEE Trans. Commu.*., vol. COM-33, no. 7, pp. 685-696, 1985.

[8] T. Eriksson, J. Lindén, and J. Skoglund, "A safety-net approach for improved exploitation of speech correlations," *in Proc. International Conference on Digital Signal Processing*, Cyprus, vol. 1, 1995, pp. 96-101.

[9] K. K. Paliwal and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, pp. 3-14, 1993.

[10] S. L. Marple, Jr., *Digital spectral analysis with applications*, Prentice-Hall, 1987.

[11] H. P. Knagenhjelm and E. Agrell, "The Hadamard transform - a tool for index assignment," *IEEE Trans. Inform. Theory*, vol. 42, no. 4, pp. 1139-1151, 1996.

[12] J. A. Rice, *Mathematical statistics and data analysis*, Wadsworth & Brooks/Cole, 1988.