Proc. Biennial Symp. Commun. (Kingston, ON), pp. 312-315, June 1990

# COMPARISON OF LPC COEFFICIENT QUANTIZERS

John Grass [1] and Peter Kabal [1,2]

[1] Electrical Engineering
McGill University
Montréal, Québec,
Canada, H3A 2A7

[2] INRS-Télécommunications
Université du Québec
Verdun, Québec,
Canada, H3E 1H6

## Abstract

Experimental results of the quantization of Linear Predictive Coded (LPC) coefficients using two general approaches, scalar coefficient quantization and vector quantization, are presented. The LPC coefficients were quantized in several domains: Line Spectral Frequency (LSF), cepstral, predictor, reflection and autocorrelation. Two distortion measures were used to evaluate the quantizers; Itakura-Saito and RMS log spectral distortion measure. The vector quantizers showed good results for only 9 bits per frame of 150 speech samples.

## 1. Introduction

Considerable investigation has been carried out into the use of Linear Predictive Coded (LPC) coefficients for the coding of speech because they provide an accurate and economical representation of relevant speech parameters. For low bit rate speech coders in particular, LPC has proven to be a popular technique.

The first step of a speech coder using LPC coefficients is to divide the discrete input speech in to segments of 10 to 30ms. An analysis of this data is performed to produce the LPC coefficients for the frame of data points. After the LPC coefficients have been obtained, the next step is to filter the speech input using the inverse filter $A(z)$ determined from the LPC coefficients. The transmitting of the LPC coefficients and the residual speech from the inverse filter are two separate speech coding tasks. The goal is to have both signals reproduced as faithfully as possible at the receiver so that the original speech signal can be reproduced by filtering the residual speech with the LPC coefficients. The different nature of the residual signal and the LPC coefficients result in very different strategies in coding the two. The methods of coding the LPC coefficients will be considered in this work. The diagram of the simulation model for studying the coding of LPC coefficients is shown in Fig. 1.

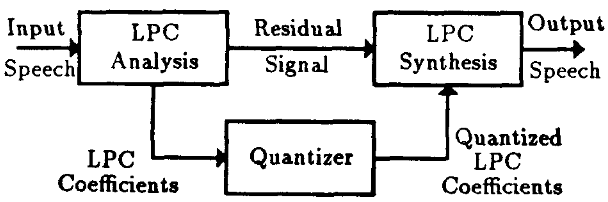


**Fig. 1** Quantization of LPC coefficients

Since only the coding of LPC coefficients is investigated in this work, the residual signal is transmitted directly to the receiver without any coding and thus without any degradation. The input speech was high quality recorded, sampled at a rate of 8 kHz. The LPC Analysis block determines the LPC coefficients and inverse filters the input speech signal to produce the residual signal. The Quantizer block codes the LPC coefficients using quantization techniques. The quantized LPC coefficients are transmitted to the LPC Synthesis block which uses them to filter the residual signal and produce the output speech. The inverse FIR filter $A(z)$ should be minimum phase to ensure all its zeroes will be within the unit circle in the $z$-domain. This guarantees that the LPC synthesis filter $\frac{1}{A(z)}$ will have all its poles within the unit circle and hence be stable. The effectiveness of the LPC coefficient quantizer is evaluated by comparing the quality of the output speech to the input speech. Comparisons are made by listening to the speech as well as using quantitative measures that compare the accuracy of the spectrum of the output speech to that of the input speech.

There are two basic approaches to quantizing the LPC coefficients. The first, scalar quantization, quantizes the LPC coefficients individually while the second approach, vector quantization, quantizes the LPC coefficients as a vector. Each approach shall be examined to investigate their performance in speech coding.

## 2. Distortion Measures

Comparisons between the received waveform of the speech coder and its transmitted waveform are made to evaluate its performance. The sound quality of a given speech coder is a perceptual measure and hence difficulties arise in finding a quantitative measure to use to accurately compare speech coders. The spectrums to be compared can be modelled in the $z$-domain as follows;

$$S(z) = \frac{g}{A(z)}\,, \qquad S'(z) = \frac{g'}{A'(z)}\,.$$

Two perceptual distortion methods that will be examined here are the RMS log spectral measure and the Itakura-Saito measure.

The RMS log spectral distortion measure is defined by the equation

$$\mathrm{RLS_{DIF}} = \int_{-\pi}^{\pi} |V(\theta)|^2 \frac{d\theta}{2\pi}$$

$$\text{where} \quad V(\theta) = \ln\left[\frac{g^2}{|A(e^{j\theta})|^2}\right] - \ln\left[\frac{g'^2}{|A'(e^{j\theta})|^2}\right].$$

One efficient method of implementing the RMS log spectral measure is to use the cepstral coefficients of the speech spectrum [1]. The cepstral coefficients for the spectrum can be calculated directly from the predictor coefficients. The cepstral coefficients are defined from the Taylor series expansion

$$\ln[A(z)] = -\sum_{k=1}^{\infty} c_k z^{-k}.$$

Although an infinite number of cepstral coefficients result from the predictor coefficients, it has been found that limiting the cepstral coefficients to three times the number of predictor coefficients is sufficient to calculate an accurate distortion measure [1].

Once the cepstral coefficients are obtained, the RMS log spectral distortion measure is simply calculated as follows;

$$\mathrm{RLS_{dB}} = \frac{10}{\ln(10)}\sqrt{2\sum_{i=1}^{N_{cep}}(c_i - c_i')^2}.$$

The Itakura-Saito measure generally corresponds closer to the perceptual quality of speech than does the RMS log spectral measure [2]. Hence, its use to evaluate the performance of speech coders is valuable. The Itakura-Saito maximum likelihood spectral distance is given by [1]

$$\mathrm{IS_{DIF}} = \int_{-\pi}^{\pi}\left[e^{V(\theta)} - V(\theta) - 1\right]\frac{d\theta}{2\pi}.$$

After simplification and for equal gains ;

$$\mathrm{IS_{dB}} = 10\log\left[\frac{\delta}{\alpha} - 1\right].$$

The residual energy $\delta$ results from passing the original signal through the filter $A'(z)$ while the residual energy $\alpha$ is obtained by passing the original signal through the filter $A(z)$. To calculate the residual energies $\delta$ and $\alpha$, the predictor coefficients of each spectrum are required.

## 3. Scalar Quantized LPC Coefficients

Most LPC coefficient quantizers use scalar quantization which quantizes each LPC coefficient independently from the other coefficients. The representation of the LPC coefficients used for coding plays an important role in the quantization process. Some representations lend themselves better to quantization than others by being less sensitive to quantization errors that affect the constructing of the LPC filter. It has been shown that it is easier to quantize reflection coefficients than predictor coefficients [3]. Another LPC coefficient representation that has good quantization properties is the Line Spectral Frequencies (LSF's) [4]. They have shown to be related to the formant frequencies in speech. Reflection coefficients have been the most popular representation of the LPC coefficients for scalar quantization in past years while considerable recent work has focused on the use of LSF's. Scalar quantization of the reflection coefficients and LSF's is investigated below.

In the reflection domain, three quantizers were tested; LPC-K8P, LPC-K10P and LPC-KGH. They each quantized the reflection coefficients separately using predetermined tables. LPC-K8P quantizes 8 reflection coefficients with 43 bits while LPC-K10P and LPC-KGH quantize 10 reflection coefficients. LPC-K10P uses 40 bits while LPC-KGH uses 24 bits. All the quantizers are non-uniform in spacing in the reflection domain, using uniform spacing in the log-area ratio domain and use a frame size of 150 samples with an overlap of 25 samples. The number of quantizing levels of each coefficient are chosen to achieve good perceptual results. The levels for LPC-KGH are taken from reference [5].

Performance of the quantizers are shown in Table 1. Two distortion methods are used for evaluation, RMS log spectral distortion measure (RLS) and Itakura-Saito distortion measure (IS). The average values of the distortion measures for four sentences, three English (CATF, CATM and PROM) and one French (PB1M).

| IS | CATF | CATM | PROM | PB1M | Ave |
|---|---|---|---|---|---|
| LPC-K8P | 0.136 | 0.083 | 0.165 | 0.937 | 0.330 |
| LPC-K10P | 0.032 | 0.029 | 0.033 | 0.176 | 0.068 |
| LPC-KGH | 0.444 | 0.399 | 2.325 | 0.635 | 0.951 |
| **RLS** | | | | | |
| LPC-K8P | 1.556 | 1.009 | 1.155 | 1.984 | 1.426 |
| LPC-K10P | 1.001 | 0.965 | 0.982 | 1.347 | 1.074 |
| LPC-KGH | 2.676 | 2.677 | 3.235 | 2.895 | 2.871 |

**Table 1** Performance of Reflection Quantizers

The first quantizers, LPC-K8P and LPC-K10P, are similar in bit size per frame but LPC-K10P outperforms LPC-K8P significantly. This shows the effects of the number of coefficients and the assignment of bits on the performance. LPC-KGH, with 24 bits per frame, performed well for the small number of bits it uses.

LSF's were first introduced by Itakura [6] in 1975. Their use as LPC coefficient coding parameters is useful due to the direct relationship between the LSF's and the formant frequencies. For example, higher order line-spectra need only be quantized coarsely since they have low perceptual impact on the quality of speech.

Two different approaches were taken to code the LSF coefficients. Both methods consider taking the coefficients as belonging to pairs. The first quantizers, Kang and Fransen design [7], calculate the center frequency of each pair and the difference or offset between the pair members.

$$\text{Center} = \frac{l_i + l_{i+1}}{2}, \quad \text{Offset} = \frac{l_{i+1} - l_i}{2}, \quad i = 1,3,5,7,9.$$

The second quantizers, based on Crosmer and Barnwell design [8], quantize the even frequencies from their relative positioning to the neighboring odd frequency LSF's. The odd frequencies are quantized with differential pulse code modulation (DPCM). The equations for the LSF coefficients are

$$\text{Even} = \frac{l_i - \hat{l}_{i-1}}{\hat{l}_{i+1} - \hat{l}_{i-1}}, \quad \text{Odd} = l_i - \hat{l}_i', \quad i = 1,3,5,7,9,$$

where $\hat{l}_i$ is a quantized LSF coefficient and $\hat{l}_i'$ is a quantized LSF coefficient from the previous frame.

To guarantee a stable synthesis filter, the LSF's should be in ascending numerical order. Due to the quantization methods, the LSF's occasionally end up crossing over. Thus, after unquantizing the LSF's, cross-overs must be checked for and corrected. The better the quantization, the fewer cross-overs should occur. The cross-overs were corrected by changing the positioning of the LSF coefficients until they were in ascending order.

Three quantizers based on the center/offset principle were examined, The first, LSF-CO21K, uses Kang and Fransen's table for the quantization frequencies. The second, LSF-CO21, uses a design based on the statistical distribution of the center and offset frequencies [9]. Improvement of this second quantizer was attempted in the third quantizer, LSF-CO30, by increasing the number of bits by about fifty percent.

Two quantizers were based on the odd/even principle; LSF-EO21 and LSF-EO30. The first uses 21 bits while the second uses 30 bits. All the LSF quantizers used a frame size of 150 samples and an overlap of 25 samples. Results of the LSF quantizers are shown in Table 2, using the Itakura-Saito distortion measure (IS) and RMS log spectral measure (RLS) with the average values calculated.

| IS | CATF | CATM | PROM | PB1M | Ave |
|---|---|---|---|---|---|
| LSF-CO21K | 32.354 | 32.212 | 47.849 | 27.870 | 35.071 |
| LSF-CO21 | 0.383 | 0.385 | 0.416 | 0.371 | 0.389 |
| LSF-CO30 | 0.099 | 0.114 | 0.237 | 0.089 | 0.135 |
| LSF-EO21 | 0.287 | 0.276 | 0.245 | 0.201 | 0.252 |
| LSF-EO30 | 0.078 | 0.094 | 0.096 | 0.091 | 0.090 |
| RLS | | | | | |
| LSF-CO21K | 9.412 | 9.470 | 9.650 | 8.646 | 9.295 |
| LSF-CO21 | 3.059 | 3.126 | 3.087 | 3.256 | 3.132 |
| LSF-CO30 | 1.648 | 1.677 | 2.074 | 1.534 | 1.733 |
| LSF-EO21 | 2.509 | 2.509 | 2.320 | 2.308 | 2.411 |
| LSF-EO30 | 1.496 | 1.557 | 1.637 | 1.663 | 1.588 |

**Table 2** Performance of LSF Quantizers

Results of the statistically determined LPC-CO21 quantizer were better than those of the LPC-CO21K quantizer. The problem with the LPC-CO21K quantizer is its lowest possible offset frequency is 300 Hz. Hence the minimum distance between LSF's in a pair is 600 Hz. This constraint was the prime reason for the poor performance of LPC-CO21K, particularly with quantizing the lower LSF's. The nine extra bits in LPC-CO30 resulted in further improvement of the LPC-CO21 quantizer. In comparison to the reflection coefficient quantizers, these quantizers offer better results. The quantizers using the even/odd principle take advantage of frame-to-frame correlation by quantizing odd frequencies using DPCM.

## 4. Vector Quantized LPC Coefficients

Vector quantization considers the set of LPC coefficients of one frame of speech input as a block. The goal of vector quantization is to remove interparameter correlation and hence reduce the number of bits required to send the set of LPC coefficients. The basic idea is to compare as a vector the LPC coefficients from a frame of speech input to pre-determined coefficient vectors stored in a codebook. The index of the closest vector is transmitted. For example, if a LPC coefficient vector is found to be closest to vector 64 in the codebook, the number 64 is transmitted to the receiver. The receiver then uses vector 64 in its copy of the codebook for the LPC synthesis of speech.

Three issues that effect the accuracy of vector quantization are the size of the codebook, composition of the codebook and the method of determining the distance between vectors. The larger the codebook, the better chance of a given vector being represented. The size of the codebook is limited by the number of bits allotted for transmission of the LPC coefficients. The method for developing the composition of the codebook used in this work is the popular Linde Buzo Gray (LBG) algorithm [10]. This algorithm takes a large number of vectors (at least several times larger than the size of the codebook to be constructed) and obtains from them a set of vectors that best represents the data vectors. This set comprises the codebook.

The same LPC coefficients in different representations will result in different vector distances from the measurement formula $(\mathbf{x}-\mathbf{y})^2$ and hence a different composition of the codebook. As shall be seen, the performance of the codebooks is affected by the choice of LPC coefficients domain to construct the codebook.

The third issue mentioned that effects the vector quantization is the distance measurement between vectors used for selecting the vectors from the codebook when quantizing. In the LBG algorithm described above, the Euclidean vector distance was used. Other possibilities include the Itakura-Saito distortion measure which generally corresponds better to the perceptional quality of speech [2]. Further, weighting schemes can be used to key on specific coefficients that have increased perceptual importance. The weighting schemes would depend on which representation of the LPC coefficients are used.

During the development of the vector quantization codebooks, LPC coefficient vectors could result that would produce an unstable filter. The reason why unstable vectors can result from a set of stable vectors is due to the method of splitting of the centroids in the LBG algorithm. Unstable vectors must be removed from the codebook or corrected to ensure that a set stable filter coefficients will be sent to the receiver.

The performance of the VQ codebooks were compared using the two distortion measures; Itakura-Saito (IS) and RMS log spectral (RLS). Codebooks in six LPC coefficient domains were tested; VQ-A9 - Autocorrelation, VQ-L9 - Line Spectral Frequency, VQ-X9 - Roots of Chebyshev polynomial expansion, VQ-C9 - Cepstral, VQ-P9 - Predictor, and VQ-K9 - Reflection. The roots of the Chebyshev polynomial expansion is a direct transformation of the LSF coefficients, using the transform $x = \cos(w)$ where $w$ is the frequency of a LSF. With 9 bits, the quantizers have a 512 level codebook and use a frame size of 150 speech samples and an overlap of 25 speech samples. The quantizers use an exhaustive codebook search method. The results are shown in Table 3.

| IS | CATF | CATM | PROM | PB1M | Ave |
|---|---|---|---|---|---|
| VQ-A9 | 1.131 | 1.773 | 2.274 | 2.939 | 2.029 |
| VQ-L9 | 0.242 | 0.310 | 0.593 | 0.981 | 0.532 |
| VQ-X9 | 0.357 | 0.429 | 0.624 | 2.543 | 0.988 |
| VQ-C9 | 0.250 | 0.265 | 0.576 | 0.738 | 0.457 |
| VQ-P9 | 0.598 | 0.968 | 1.164 | 3.588 | 1.580 |
| VQ-K9 | 0.512 | 0.673 | 1.008 | 2.632 | 1.206 |
| RLS | | | | | |
| VQ-A9 | 4.826 | 4.925 | 6.133 | 6.617 | 5.625 |
| VQ-L9 | 2.535 | 2.748 | 3.536 | 4.043 | 3.216 |
| VQ-X9 | 2.818 | 2.962 | 3.626 | 4.559 | 3.491 |
| VQ-C9 | 2.518 | 2.630 | 3.473 | 3.831 | 3.113 |
| VQ-P9 | 3.420 | 3.529 | 4.230 | 5.084 | 4.066 |
| VQ-K9 | 3.190 | 3.317 | 4.352 | 5.375 | 4.059 |

**Table 3** Performance of Vector Quantizers

It should be noted that the codebooks were trained using only English sentences. The French sentence, PB1M, was used to test the flexibility and robustness of the quantizers. The French sentence shows how the vector quantizers are restricted by the scope of their training sentences while the individual coefficient quantizers are more flexible. The general trend for the quantizers to perform better on the English sentence CATF than on the English sentence PROM was undoubtedly not a coincidence. The training data had the same speaker as the one speaking the CATF8 sentence and not the speaker uttering the PROM8 sentence.

The codebooks VQ-A9, VQ-P9 and VQ-K9 did not perform very well. The codebook VQ-X9 was quite good for the English sentences but was very poor on the French sentence. The two quantizers VQ-L9 and VQ-C9 performed very well overall. These quantizers, with only 9 bits, outperformed the scalar reflection coefficient quantizer, VQ-KGH, which uses 24 bits.

Although one set of LPC coefficients can be represented in any of the representations used above in the VQ quantizers, the quantizers did not perform the same despite having the same training data for the codebooks, method of constructing the codebook and the method of searching the codebook. The reason for the differing performances is due to the method of determining which vector is considered closest in distance to the vector to be quantized. The Euclidean distance $(\mathbf{x}-\mathbf{y})^2$ has different meaning for the different representations of the LPC coefficients. By examining the distortion measure graphs for the quantizers VQ-K9 as an example, it is seen that it performs poorly throughout with some very large error spikes. In particular, in frames 69 to 70, the VQ-K9 goes from having a large error in one frame to a small one in the next frame followed by a large error in the next frame after that. The question is if there exists vectors in the codebook that would have caused less error than those chosen.

When switching the Euclidean distance in the quantizer to the Itakura-Saito distortion measure yet still using the same codebook, better vectors were chosen for the frames of speech input. In fact, the overall performance of the VQ-K9 quantizer improved to the same level as the LSF and cepstral VQ quantizers. The use of the Itakura-Saito distortion measure makes the quantizer select a vector from the codebook to more closely match the spectral envelope than when the quantizer used the Euclidean distance in the reflection domain.

From the results of the quantizer VQ-L9 and VQ-C9, we conclude that using the Euclidean distance in the LSF and cepstral domain result in good perceptual matches between the original vector and the codebook vector. The matches are good because the LSF and cepstral coefficients have a more direct relationship to the spectral envelope than do other representations of the LPC coefficients such as the reflection coefficients. The same reasoning applies to why better performance is achievable in scalar quantizing LSF coefficients rather than reflection coefficients.

## 5. Conclusion

Distortion measurements are valuable for good, quick evaluation of speech coders until more extensive subjective perceptual testing can be performed. For vector quantization, not only is a large volume of data required, the data must be balanced to give the quantizer flexibility. The vector quantizers show that for only 9 bits considerable performance can be achieved. Future areas of investigation include the implementation of interpolation schemes between frames or sub-frames of speech samples. Parameters such as codebook size, number of coefficients, frame size and overlap can be varied to try to determine the optimum parameter combination. Adding a second stage quantizer after the first vector quantization is the next step to improving the VQ coder. The second stage could be a scalar coder or another VQ coder. Investigation is now being performed into the second stage quantizer.

## References


1. A. H. Gray, Jr. and J. D. Markel, "Distance measures for speech processing", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-24, pp. 380–391, Oct. 1976.
2. R. M. Gray, A. Buzo, A. H. Gray, Jr. and Y. Matsuyama, "Distortion measures for speech processing", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-28, pp. 367–376, Aug. 1980.
3. D. O'Shaughnessy, *Speech Communications*, Addison-Wesley, Don Mills, Ontario, 1987.
4. G. S. Kang and L. S. Fransen, "Application of Line Spectrum Pairs to Low Bit Rate Speech Encoders", *Proc. Int. Conf. on Acoust. Speech, Signal Processing*, Tampa Florida, pp. 7.3.1–7.3.4, April 1985
5. O. Ghitza and J. L. Goldstein, "Scalar LPC quantization based on formant JND's", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-34, pp. 697–708, Aug. 1986.
6. F. Itakura, "Line Spectrum Representation of Linear Predictor Coefficients of Speech Signals", *J. Acoust. Soc. Amer.*, vol. 57, S35(A), 1975.
7. G. S. Kang and L. J. Fransen, "Low-bit rate speech encoders based on line-spectrum frequencies (LSFs)", *Naval Research Laboratory Report 8857*, Nov. 1984.
8. J. R. Crosmer and T. P. Barnwell, III, "A Low Bit Rate Segment Vocoder Based on Line Spectrum Pairs", *Proc. Int. Conf. on Acoust. Speech, Signal Processing*, Tampa Florida, pp. 7.2.1–7.2.4, April 1985.
9. C. C. Chu and P. Kabal, "Coding of LPC Parameters for low bit rate speech coders", *Rapport Technique de l'INRS-Télécommunications No. 87-19*, March 1987.
10. Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantization design", *IEEE Trans. on Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.