THE EFFECT OF LPC ORDER ON THE PERFORMANCE OF VECTOR QUANTIZATION IN ISOLATED-WORD RECOGNITION

W. A. Mahmoud and L. A. M. Bennett

Department of Electrical & Electronic Engineering
University College of Swansea
Singleton Park
Swansea SA2 8PP
United Kingdom

1 ABSTRACT

The Vector quantization (VQ) of LPC spectra has been applied to cocoding and also very recently to speech recognition as a means of reducing memory requirements for the storage of reference templates and of reducing computation time. This paper examines the effect of the LPC order (P) on the distortion measure and on the performance of the VQ algorithm in isolated-word speech recognition (IWSR).

2 INTRODUCTION

Linear predictive coding (LPC) coefficients have become the most powerful and predominant features for representing the speech signal. The number of LPC spectra required to describe the words of a vocabulary is very high. The basic concept of VQ is to classify these LPC spectra by comparing them with vectors in a codebook. The goal of a VQ algorithm is to minimise the distortion measure associated with the classification procedure.

Several factors that effect the distortion have been studied including the initial codebook, the multiplying factors type of distance measure etc. In this paper the effect of an important factor, the LPC order P, on the distortion as well as on the performance of VQ in IWSR is considered. A speech recognition system has been developed in software on a 68K mini-computer using Dynamic Time Warping (DTW) and Vector Quantization (VQ). The recognition error rates against codebook sizes of 4,...,128 codewords have been obtained and compared for five different values of LPC order P.

3 THE VQ ALGORITHM

Assume that a training set of V vectors in the form of gain normalised autocorrelation terms is given. It is desired to find a codebook of size C codewords such that the average distortion measure (distance) (DS (C)) of a vector in the training set from the closest codeword is minimised, thus:

$$DS(C) = \underset{(C)}{\text{Min}} \left| \frac{1}{V} \cdot \sum_{i=1}^{V} \underset{1<m<C}{\text{Min}}\, d(v_i, c_m) \right| \quad [1]$$

where $d(v_i, c_m)$ is the LPC distance between the training vector v_i and the codeword c_m. The log likelihood distance measure of Itakura is used.

4 THE EXPERIMENTAL BACKGROUND

Seven speakers, five male and two female, generated the database of spoken Arabic digits (0-9). In one session each speaker was asked to contribute ten digits as isolated utterances. In a second session each speaker was asked to read out as isolated-words a list of one hundred digits in random order. The first training set was used to generate the codebook of sizes of 2, 4,....,128 codewords using the above vector quantization algorithm for five different values of P. A fourth-order antialiasing elliptic filter with cut-off of 4.8 kHz was used together with 12-bit ADC and a sampling rate of 10 kHz.

5 THE EFFECT OF LPC ORDER P ON THE DISTORTION MEASURE OF VQ

To obtain useful results with vector quantization it is important to understand the relationships and the effects of the choice of order of the LPC model on the distortion measure.

To evaluate the effects of P on the performance of the VQ algorithm, a series of two sets of tests were run. These series of experiments consisted of the design of the VQ for the Arabic-digit vocabulary for two Hamming window lengths, 12.8 msec and 19.2 msec. Five different values of P were used, which were, 8, 10, 12, 14 and 16. Fig. 1 shows the distortion measure of VQ for these values of P as a function of codebook size for the 12.8 msec Hamming window. A similar result was achieved for the 19.2 msec window. It is clear from these plots that the distortion measured increases as P increases. This is understandable, because when the value of P increases, more of the details of the spectrum are included in the LPC spectrum.

This was observed from the plots of their LPC spectra where the 8 and 10-pole models give much smoother spectra than the 12 to 16-pole models for a given frame of speech. Therefore the distance measure between two LPC spectra for smaller values of P will be smaller than that between two spectra of higher P.

The natural question now is how many poles should one use in fitting the model for the data acquisition system under consideration? There is no direct answer to the above question as far as the distortion measure is concerned. Therefore to understand better the effect of P on the VQ algorithm it is necessary to study its performance outside the training algorithm.

6 THE EFFECT OF P ON THE PERFORMANCE OF VQ IN IWSR

To evaluate further the effect of P on the performance of the VQ outside the training data a series of recognition tests were carried out. Two sets of experiments were performed on the DTW/VQ and LPC/DTW isolated word recognizers.

(a) The first set of experiments was performed for the Hamming window of 12.8 msec length. The speaker-dependent mode of recognition was used, hence each speaker was treated separately. The 100 digits of each speaker were used as templates once and as tests next, hence a total of 10,000 crossing were performed for each speaker.

(b) In the second test the above experiments were repeated for a Hamming window of 19.2 msec.

The templates and the tests were quantized to a VQ codebook of size of 4,8, ..., 128 codewords and the recognition performance was compared. In this way some direct results are obtained for recognition error rate againsts VQ codebook size. The results of the first tests are given in Fig. 2, which shows plots of error rate versus codebook size. The results of the second test are given in Fig. 3.

From inspection of these plots it is clear that the 8 and 10-pole models provided insufficient recognition accuracy. For the 12 to 16-pole models the recognition error rate was acceptable and there was a slight difference in the recognition accuracy for them, particularly for codebook sizes of 32 and more. As a practical matter, it is generally desirable to use the minimum number of poles necessary to model accurately the significant features of the signal. Therefore, it was decided that the 12-pole model was sufficient for the recognizer under consideration.

7 CONCLUSIONS

This paper has studied the effect of P on the distortion measure and performance of VQ in IWSR. Two sets of experiments were run on a database of 700 isolated digits from 7 speakers. Increasing P increases the VQ distortion, however it improves its performance outside the training data. The results strongly suggest that, for the VQ recognizer under consideration, the minimum, value of P should be 12. This reinforces the result reported elsewhere for the DTW recognition systems, that the minimum P should be 2 poles for each kHz band of the filter plus two extra poles.

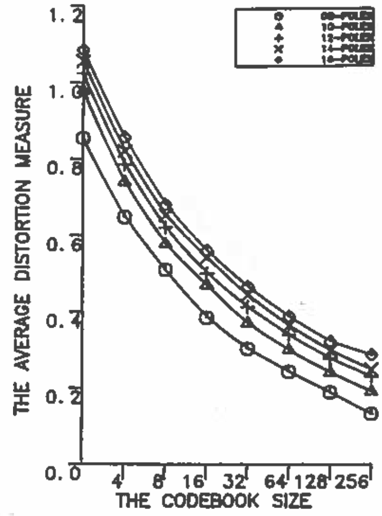

Fig.(1) The distortion measure for five LPC order, for the 12.8 msec window, as a function of codebook size.

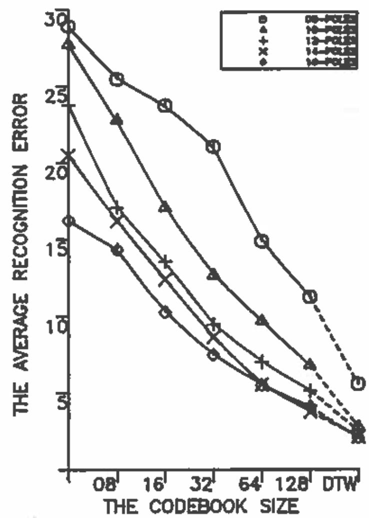

Fig.(2) The Average Recognition Error using a window of 12.8 msec, for five values of P as a function of codebook size.

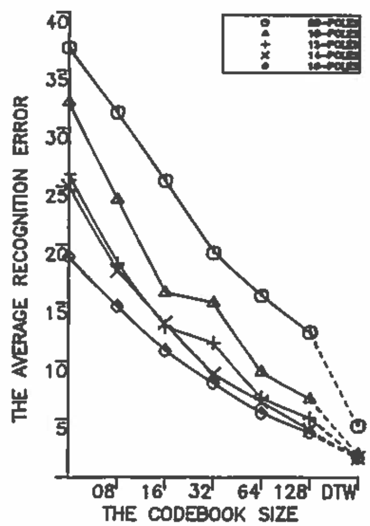

Fig.(3) The Average Recognition Error using a window of 19.2 msec, for five values of P as a function of codebook size.