CODING OF THE LPC SPECTRAL PARAMETERS
USING VECTOR PREDICTIVE QUANTIZATION

D. J. Zarkadis and B. G. Evans
Department of Electronic and Electrical Engineering
UNIVERSITY OF SURREY
Guildford GU2 5XH, Surrey, U.K.

Abstract

The LPC-type speech coders require efficient quantization and coding of the LPC spectral parameters in order to operate at low bit-rates. We present here results from an investigation to the Vector Predictive Quantization (VPQ) approach for the compression of the LPC spectral information. Reduction of the computation requirements of the VPQ scheme by means of Split-Vector Quantization is considered. The distortion-rate characteristic of the coding scheme incorporating the Split-Vector Quantizer is estimated, based on log-spectral distortion measures. Our results indicate that this coding scheme can operate at 20 to 22 bits/frame without any perceptually significant distortions in the synthesized speech.

1. Introduction

There has been considerable interest on the coding of the LPC spectral parameters and various methods have been proposed and investigated [1]-[5]. The benefits from the compression of the LPC spectral information can be twofold for the LPC-based speech coders: either at a given bit-rate the saved bits can be used to improve the quantization of the excitation information of the speech coders enhancing their performance, or lower bit-rates can be achieved.

Vector Quantization [6][7] has been widely used in many data compression systems. In this paper we investigate the Vector Predictive Quantization approach [1] for compressing the LPC spectral information. The spectral envelope of the short-time speech spectrum changes slowly from frame to frame and significant correlations between successive speech spectra exist. Then, the VPQ scheme can achieve bit-rate reductions by applying Vector Prediction on the spectral envelope vectors and subsequently Vector Quantization on the residual vectors. Reduction in the computation requirements of the original scheme is achieved by introducing a Split-Vector Quantizer (SVQ) and the combined scheme is denoted here as VP-SVQ.

Performance indicators in the log-spectral domain are considered and the distortion-rate characteristic of this coding scheme is estimated.

2. Spectral representation of the LPC Information

The procedures for the representation of the LPC information at the input of the VP-SVQ scheme, as well as the synthesis operation for the recovery of this information at the receiving end, are similar to that given in [1]. An m-th order LPC analysis (autocorrelation method) is applied every 20msec, on 25msec windowed and asymmetrical overlapping frames of speech. The LPC spectral information is derived from the autocorrelation set of the impulse response of the inverse filter \( \mathbf{A}_n(z) \). This results to an \((M+1)\)-dimensional vector \( \mathbf{X}_n = (x_{n,0}, x_{n,1}, \ldots, x_{n,M}) \) which corresponds to a sampled version of the spectral envelope for the short-term speech spectrum for the n-th frame. For the synthesis operation we use the autocorrelation set derived from the quantized spectral envelope vector \( \mathbf{X}_n \).

We set \( m = 10 \) and \( M = 30 \) as the parameters of the analysis/synthesis procedure.

3. Description of the VP-SVQ scheme

In this scheme the vector predictor exploits the interframe correlations of the spectral envelope vectors and then the Split-Vector Quantizer takes care of the dependencies within the residual vector components. A block diagram of the VP-SVQ scheme is shown in figure 1. Its operation is based on a two-step procedure. During the first step the VP-SVQ scheme predicts the input spectral envelope \( \mathbf{X}_n \) from past-quantized spectra using a predictor codebook. At the second step the residual spectrum is quantized using the Split-Vector Quantizer (SVQ). The coding rate of the scheme is associated with the transmission of the codeword indices of the codebooks involved in this process.

3.1 The Vector Predictor

A codebook of \( B_p \) bits is associated with the vector predictor. This codebook contains the predictor coefficients vectors \( \mathbf{C}_i = (c_{i,1}, c_{i,2}, \ldots, c_{i,p}) \) for \( i = 1, 2, \ldots, 2^{B_p} \) and \( p \) the order of the predictor. The design of the predictor codebook was based on a procedure similar to the unweighted version of [1] using a training sequence of 24000 vectors. During the actual operation, all the possible estimates \( \tilde{X} \) for the input vector \( \mathbf{X}_n \) are formed, as a linear combination of past quantized spectra \( \mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \ldots, \mathbf{X}_{n-p} \) using the predictor coefficient vectors \( \mathbf{C}_i \) from the predictor codebook. Then the optimum predictor coefficient vector for the n-th frame is chosen such that the squared-error

\[
\text{d} (X_n, \tilde{X}) = (X_n - \tilde{X})^T (X_n - \tilde{X})
\]

is minimized. The residual envelope vector \( R_n \) is formed as a result of the prediction process.
3.2 The Split-Vector Quantizer

The residual vector $R_n$ is split into two subvectors $R_{1,n}$ and $R_{2,n}$ as shown in figure 1, of dimensions 16 and 15 respectively. The subvectors are then fed into the vector quantizers and are quantized independently. Each vector quantizer is associated with its own codebook $C_1$ or $C_2$ or $B_1$ and $B_2$ bits respectively. Here, we set $B_1 = B_2 = Br/2$ with $Br$ being the number of bits for a full-VQ codebook. The codebooks were designed using the LBG binary splitting technique [8]. During the actual operation, the quantized subvectors $\hat{R}_{1,n}$ and $\hat{R}_{2,n}$ are chosen such that the squared-error distortion

$$d(R_{i,n}, \hat{R}_{i,n}) = (R_{i,n} - \hat{R}_{i,n})^T(R_{i,n} - \hat{R}_{i,n})$$

is minimized. The quantized subvectors are then combined in order to form the $\hat{R}_n$ vector and close the system loop. The sequence $\{X_n\}$ of the spectral envelope vectors represents the variations of a gain-normalized all-pole filter. Therefore, the residual vectors contain only shape information and no extra gain parameters need to be transmitted. Splitting of the residual envelope vectors was also reported in [9].

The use of SVQ reduces the computation requirements of the scheme in comparison to a full-VQ case. More specifically, the search of a codebook of $2^{Br/2}$ 31-dimensional vectors for the full-VQ is substituted by searching two codebooks of $2^{Br/2}$ vectors each, of dimensions 16 and 15. The storage requirements are also reduced. The use of the SVQ was also motivated here by the fact that the vector predictor decorrelates the components of the spectral envelope vectors. This means that the components of the residual subvectors are associated with reduced interdependencies and therefore, any losses from splitting and independent quantization should also be reduced.

4. Experimental results

The functional elements of this predictive coding scheme are the Vector Predictor and the Split-Vector Quantizer. Their combined effect determines the performance of the overall scheme, and we used the prediction gain and the quantizer-SNR as indicators of their performance. In order to assess the overall performance of the coding scheme, we used a spectral distortion measure which accounts also for the distortions introduced by the analysis/synthesis process. A database of 2000 frames of speech (remained after silence elimination) including 2 male and 2 female speakers was used as the test sequence. The performance of the VP-SVQ scheme was tested at rates between 14 and 26 bits/frame with bit allocations according to Table 1.

| Predictor codebook (bits) | (0,1,2, ..., 8) | 8 | 8 |
| Residual codebooks (bits) | (7, 7) | (8, 8) | (9, 9) |
| Total bits/frame | (14,15, ..., 22) | 24 | 26 |

The Vector Predictor

The prediction gain of the scheme is defined in the log-spectral domain as

$$PG = \frac{\sum_{n} \|X_n\|^2}{\sum_{n} \|R_n\|^2}$$

where $\|\cdot\|$ denotes the Euclidean norm. We estimated the prediction gain as function of the predictor order and the codebook size. We noticed that the difference in the performance of the vector predictor as function of its order is more noticeable at large size codebooks, and that the prediction gain for the 1st and 2nd order predictors saturates at small size codebooks. We have chosen a 4th order predictor ($p = 4$) for the rest of our experiments, which is associated with prediction gains in the range of 7.2 to 10.5 dB for $Bp = 0$ to 8 bits. This is due to the high frame to frame correlations of the spectral envelope vectors. Figure 2 shows the histogram for the occurrence of the predictor codebook vectors for the case of a 1st-order predictor and a 3-bit codebook (6 codevectors). In each case, each codevector is reduced to a single coefficient $c_{i,1}$ and the interpretation of 1st-order autocorrelation coefficient can be given to it. Then, the frequent occurrence of coefficients associated with the high-correlation zone (0.8, 1.1) indicates the significant correlation between successive spectra and explains the prediction gains obtained.

![Figure 2. Histogram for the occurrence of the predictor codebook coefficients ($Bp = 3$ bits and $p = 1$ case).](image-url)
The Split-Vector Quantizer

The performance of the Split-Vector Quantizer is determined by the quantizer-SNR, defined in the log-spectral domain as

\[
QSNR = \frac{\sum \| R_n \|^2}{\sum \| R_n - \hat{R}_n \|^2} \tag{4}
\]

In order to assess the effects of splitting on the performance of the scheme, the (7,7)-bit SVQ was compared to a 14-bit full-VQ case. For the full-VQ we used a random codebook of 2^14 vectors of dimension 31. The codevectors were selected from a residual sequence \([R_n]\) obtained by operating the vector predictor in an open-loop mode. At high rates, the performance of random codebooks converges asymptotically to that of the structured codebooks [10], therefore, it seems sensible here to use it as reference for the full-VQ case. Results for the Vector Predictive Coding scheme with the full-VQ and SVQ, operating in open and closed-loop modes are given in Table 2. The log-spectral distortions estimated according to equations (5) and (6) have also been included.

Table 2.

Performance of the Vector Predictive coding scheme with full-VQ and SVQ, in open and closed-loop modes (20 bits/frame, \(p = 4\))

<table>
<thead>
<tr>
<th>Performance indicators</th>
<th>Vector Predictor with full-VQ</th>
<th>Vector Predictor with SVQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>PG (dB)</td>
<td>o-1: 10.76</td>
<td>c-1: 10.76</td>
</tr>
<tr>
<td>QSNR (dB)</td>
<td>o-1: 4.68</td>
<td>c-1: 4.30</td>
</tr>
<tr>
<td>Dav (dB)</td>
<td>o-1: 1.87</td>
<td>c-1: 2.05</td>
</tr>
</tbody>
</table>

We note that the SVQ is associated with lower quantizer-SNR's than the full-VQ, both at the open-loop and closed-loop modes. This is due to the remaining interdependencies between the components of the two subvectors as well as to the reduced dimensionality of the subvectors. According to our results, the closed-loop mode does not amplify significantly the losses induced by the splitting and the overall losses are rather limited. For this specific case, the reduction in the computation requirements of the VQ scheme is of the order of 10\(^2\).

A sequence of spectral envelope vectors at the input of the VP-SVQ scheme are shown in figure 3a (continuous line). The reconstructed spectral envelope sequence is shown by the dotted line in figure 3a and the residual vector sequence is shown in figure 3b.

Spectral distortion

We used log-spectral distance measures [11] in order to estimate the overall spectral distortion introduced by this coding scheme. The per-frame spectral distortion \(d_n\), is estimated according to

\[
d_n = \frac{1}{2\pi} \int_0^{2\pi} \left[ 10 \log \left( \frac{A_n(\omega)}{\hat{A}_n(\omega)} \right)^2 \right] d\omega \ (dB^2) \tag{5}
\]

with \(A_n\) and \(\hat{A}_n\) being the spectra of the LPC inverse filters before and after quantization respectively for the \(n\)-th frame.

Then the rms log-spectral distortion is estimated as

\[
D_{av} = \left( \frac{1}{M} \sum d_n \right)^{1/2} \ (dB) \tag{6}
\]

The distortion-rate characteristic of the VP-SVQ scheme with the 4th-order predictor and bit-allocations according to Table 1 is shown in figure 4 on a \(dB^2\) scale. For rates of the VP-SVQ scheme in the range of 26 to 14 bits/frame the spectral distortion varies from 1.76 to 2.43 dB. At 20 bits/frame \(D_{av} = 2.05\) dB and according to Table 2, the splitting process contributes about 0.18 dB to the distortion. The distribution of the frame distortions (dB scale) at this rate is shown in figure 5. We believe that the distortions introduced by the VP-SVQ scheme need to be carefully balanced against the distortions in the excitation of the LPC systems. This will show the benefits of the VP-SVQ scheme. This will show the benefits of the VP-SVQ scheme for an LPC-based speech coder at the bit rate of interest.

Finally, we tried this coding scheme on the RELP speech coder described in [12], as replacement of the LPC scalar quantization scheme (log-area ratios parameters coded with 48 bits). Our experiments indicated that the VP-SVQ scheme incorporated in this speech coder can operate at 20 to 22 bits/frame, without introducing perceptually significant distortions in the synthesized speech. This corresponds to a reduction in the LPC information rate from 2.4 to 1.1 Kb/s and the bit rate of the speech coder is reduced by 1.3 Kb/s.
5. Conclusions

We have investigated issues related to the compression of the LPC spectral information based on the VPQ approach. A Split-Vector Quantizer was introduced which reduces the computation requirements of the coding scheme, and the distortion-rate characteristic of the combined scheme was estimated. Our results indicate that the VP-SVQ scheme can code the LPC spectral information at 20 to 22 bits/frame without any perceptually significant distortions in the synthesized speech.

Acknowledgement

This work was supported by the Speech and Language Processing Division of British Telecom Research Laboratories.

References