Short Paper

CLASSIFIED VECTOR QUANTIZATION OF LPC PARAMETERS

Mu-Liang Wang and Jar-Ferr Yang*

ABSTRACT

To achieve high coding efficiency, modern speech coders adopt hybrid coding approaches, which utilize different coding mechanisms for various classified speech segments. With known voiced/unvoiced detection, in this paper, a classified LPC quantization (CLPQ) scheme is presented to effectively encode line spectral frequencies (LSF). The proposed CLPQ scheme improves the performance of the classified LSF vector quantizer, which adopts two LSF codebooks derived separately from voiced and unvoiced speech frames. With an objective spectral distortion measure, the CLPQ scheme successfully reduces the bit rate by about 1 bit/frame. Many classified LSF quantizers with different codebook structures and bit rates were evaluated. It would be helpful to design a classified LSF quantizer, which arrives at a compromise between distortion, bit rate and computational complexity.

Key Words: speech coding, LPC quantization, line spectral frequency.

I. INTRODUCTION

Linear predictive coding (LPC) parameters are widely used to represent the short-term spectral envelopes of speech signals. Most speech coders adopt line spectrum frequencies (LSF) to efficiently represent LPC coefficients. The performance of LSF quantization is usually evaluated by spectral distortion (SD) measures given by

$$SD = \left( \frac{1}{\pi} \int_0^{\pi} \left| 10 \log \left( \frac{S(\omega)}{\hat{S}(\omega)} \right) \right|^2 d\omega \right)^{1/2},$$

where $S(\omega)$ and $\hat{S}(\omega)$ are the LPC power spectra obtained by the original and the quantized LSF vectors, respectively. The average spectral distortion $SD$ is the average of $SD$ over frames. To achieve transparency, the spectral quantizer must satisfy the following 3 conditions (Paliwal and Atal, 1993):

(1) The average spectral distortion $SD$ is less than 1 dB.
(2) The number of outlier frames with $SD$ in the range of 2-4 dB must be less than 1-2%.
(3) There is no outlier frame with $SD$ larger than 4 dB.

The lower boundary of the rate-distortion function for a ‘universal’ VQ has been assessed and simulation results indicate that a ‘1 dB’ average SD can be achieved at 22-23 bits/frame (Hedelin, 1994).

To encode a speech signal and achieve high quality speech at a bit rate below 4 kbit/s, an increasingly important coding approach is the use of multi-modal speech coders which apply different speech coding strategies for different classes of speech signal. A simple voicing classification strategy is used to distinguish between a voiced (V) and an unvoiced (UV) speech frame. It is reasonable to adaptively quantize the LSF parameters for specific phonetic classes of speech. The voicing-specific LPC quantization (VSLQ) scheme has been presented in (Hagen et al., 1999) and the transparency rule for the unvoiced frame has been modified as

(1) The average spectral distortion $SD$ is less than 2 dB for unvoiced frame.
(2) The percentage of outlier frames with SD above 4 dB must be less than 1%.

Moreover, Hagen et al. (1999) also predict the theoretical lower boundary of the rate-distortion function for the ‘classified’ VQ. The number of bits is 20.7 bits/frame and 10-11 bits/frame for voiced frames and unvoiced frames, respectively. However, the rate-distortion boundary for the voiced frame has not been achieved in (Hagen et al., 1999). The associatively classified partitioned VQ (ACPVQ) exploits the intra/inter subvector correlations among the LSF vector components (Chen et al., 1999). By adopting the M-L tree search algorithm, the ACPVQ achieves an average SD below 1 dB at 21 bits/frame accompanied by increases in both computational complexity and storage memory. Recently, a classified LSF quantizer with a bit budget of 21 bits has been implemented for low-rate coders (Hiwasaki et al., 2004). However, the approach also results in high computational complexity.

In this paper, we will propose a classified LSF quantizer to quantize LSF parameters by using a bit rate as low as possible. To evaluate computational complexity, coding efficiency and memory requirements, several configurations for class-specific LSF VQ have also been addressed. Thus, the proposed algorithm can help to design an efficient LPC quantizer for any hybrid speech coder to achieve the lowest feasible bit rate design.

II. SPECTRAL QUANTIZATION OF LPC PARAMETERS

In the source model, the speech signal is assumed to be an excitation signal filtered through a time-varying linear synthesis filter given by \( H(z) = 1/A(z) \), where \( A(z) \) is the optimal linear predictor represented by

\[
A(z) = \sum_{i=0}^{N_p} a_i z^{-i}.
\]

Here the set of \( \{ a_i \} \) for \( i = 0, 1, \ldots, N_p \) are the LPC coefficients. The order \( N_p \) is chosen such that the spectral envelope can be estimated adequately. It is well known that line spectral frequency (LSF) coefficients \( \{ \omega_i \}, i = 1, 2, \ldots, N_p \) provide better quantization efficiency than other representations.

Numerous quantization schemes have been explored to quantize LSF coefficients. The multi-stage vector quantization (MSVQ) scheme consists of a cascade of \( m \) VQ stages. To reduce the computational complexity, the sequential search approach is usually used. The final reconstructed LSF vector for \( m \)-MSVQ is obtained by summing the \( m \) quantized vectors of each stage:

\[
\omega_k = \sum_{i=1}^{m} \omega_k^{(i)},
\]

where \( \omega_k^{(i)} \) denotes the optimal codeword of the \( i \)th stage VQ. The M-L tree search scheme has been used to improve the performance of the MSVQ. It has been shown that the M-L tree search applied in a multi-stage procedure achieves performance very close to that of a full search (LeBlanc et al., 1993).

The split vector quantization (SVQ) is helpful to reduce the requirement for memory and computational complexity. For \( n \)-SVQ, an \( N_p \)-dimensional LSF vector \( \omega \) is divided into \( n \) subvectors. The SVQ scheme presented in (Paliwal and Atal, 1993) achieves transparency with 24 bits/frame. Besides, a predictive vector quantization (PVQ) was adopted to improve coding efficiency by considering the interframe correlation of adjacent frames. A moving average (MA) backward predictor is usually used in the PVQ. For the \( k \)th frame, the LSF parameter is predicted as

\[
\omega_{k} = \sum_{l=1}^{q} \lambda_{l}^{(k)}(k) \omega_{k-l},
\]

where \( q \) denotes the predictor order and \( \omega_{k-l} \) is the quantized LSF vector of the \( (k-l) \)th segment.

III. THE PROPOSED SCHEME

In this section, a classified LPC quantization (CLPQ) scheme is proposed. The coding procedure of the proposed CLPQ scheme is addressed in detail.

1. Proposed CLPQ Scheme

In the proposed CLPQ scheme, a predictive two-stage SVQ scheme associated with the M-L tree search method is used to quantize the classified LSF coefficients. Fig. 1 depicts the block diagram of the proposed CLPQ scheme. The V/UV decision indicates whether the voiced LSF quantizer or the unvoiced LSF quantizer is used for the current segment. The first stage VQ is the direct VQ (DVQ) without using any predictions. The second stage VQ adopts the SVQ method to reduce memory requirement and computational complexity. Beginning with the first stage codebook, the M code vectors which achieve the lowest distortion are selected and thus \( M \) associative residual vectors are calculated. The LSF residual vector can be either quantized by the DVQ or by the first-order PVQ. The one flag bit ‘p/d’ is used to indicate whether a DVQ or a PVQ is used in the second stage VQ. The second stage codebook is searched \( 2 \times M \) times. The optimal indices of codewords and flag bit ‘p/d’ are selected based on a weighted minimum square error (WMSE) criterion.
2. Distance Weighting

The LPC quantizer quantizes the LSF vector with minimum spectral distortion criterion. According to the LSF's local property, the weighted squared difference between the original and quantized LSF vector is used in the encoding process as

\[
d(\omega, \hat{\omega}) = \sum_{i=1}^{N_p} w_i (\omega_i - \hat{\omega}_i)^2,
\]

where \(\omega_i\) and \(\hat{\omega}_i\) denote the \(i^{th}\) elements of the original LSF vector and the reconstructed LSF vector, respectively. Various weighting functions have been proposed. The formant bounded weighting (FBW) function has achieved a good performance (Lee et al., 2001). Let \(\omega_{N_p,i}\) and \(\omega_{N_p-1,i}\) denote the \(i^{th}\) elements of the LSF vector of order \(N_p\) and \(N_p - 1\), respectively. The \(i^{th}\) weighting value of the FBW function is given as

![Fig. 1 Block diagram of the proposed CLPQ scheme: (a) Overall structure of the proposed scheme, (b) Detailed structure of two-stage LSF quantizer.](image-url)
\[
\begin{align*}
   w_i^\text{FBW} &= s_i^2 \left( \frac{1}{\omega_{N_p, i}^{-1} - \omega_{N_p, i-1}^{-1}} + \frac{1}{\omega_{N_p, i-1}^{-1} - \omega_{N_p, i}^{-1}} \right), \\
   i &= 1, 2, \ldots, N_p, \\
   w_i^\text{MFBW} &= s_i^2 \left( \frac{1}{\omega_{N_p, i}^{-1} - \omega_{N_p, i-1}^{-1}} + \frac{1}{\omega_{N_p, i-1}^{-1} - \omega_{N_p, i}^{-1}} \right), \\
   i &= 1, 2, \ldots, N_p.
\end{align*}
\]

where \(\omega_{N_p, i} = 0\) and \(\omega_{N_p, i} = \pi\). In this paper, the FBW weighting function is adopted to further improve the performance of the CLPQ scheme. In contrast with the FBW method, the value of each denominator in (6) is limited to not less than a given threshold. The weighting value of the modified formant bounded weighting (MFBW) function is given as

\[
   w_i^\text{MFBW} = \left( \frac{1}{\omega_{N_p, i}^{-1} - \omega_{N_p, i-1}^{-1}} + \frac{1}{\omega_{N_p, i-1}^{-1} - \omega_{N_p, i}^{-1}} \right), \\
   i &= 1, 2, \ldots, N_p.
\]

### 3. Codebook Training Procedure

The LSF codebooks were trained from voiced and unvoiced speech frames separately. The multi-stage codebooks are trained sequentially based on training with LBG algorithms (Linde et al., 1980). First, the codebook of the first stage VQ was trained with original LSF vectors based on (5). Accompanying the M-L tree search scheme, \(M\) codewords that provide the lowest weighted distortion are selected and \(M\) residual vectors for the \(k\)th frame are computed as

\[
   \omega_{2, j, k} = \omega_j - \omega_{2, j, k}^{(1)}, \quad j = 1, 2, \ldots, M.
\]

The second stage DVQ codebook is searched \(M\) times and the index, \(j^*\) corresponding to the optimal codeword of the first stage VQ is also selected. Then we have the residual vector \(\omega_{2, j^*, k}^{(2)}\) for codebook training of the second stage DVQ. The codebook of the first stage VQ and that of the second stage DVQ obtained after previous training iterations are used for the current training iteration. The iterative training process is terminated if the difference of overall weighted mean square error between the current and the previous iterations is less than a given threshold or the number of iterations is larger than a specific number.

The training procedure for the first-order PVQ is similar to the one for the DVQ. The residual vector of the \(k\)th frame, \(\omega_{2, j, k}^{(2)}\), for the second stage PVQ is given as

\[
   \omega_{2, j, k}^{(2)} = \omega_j - \lambda \omega_{2, j, k}^{(1)} - (1 - \lambda) \omega_{2, j, k-1}.
\]

The value of prediction factor \(\lambda\) is set as 0.7. The \(M\) difference error vectors, \(\{ \omega_{2, j, k}^{(2)}, j = 1, 2, \ldots, M\}\), are computed and searched. Based on the optimal index \(j^*\) corresponding to the optimal codeword of the first stage VQ, the residual vector \(\omega_{2, j, k}^{(2)}\) is obtained for codebook training of the second stage PVQ. It is noted that both the codebook of the second stage DVQ and the one for the second stage PVQ were trained based on a minimum mean square error (MMSE) criterion.

### IV. EXPERIMENTAL RESULTS

In the paper, both the training and test sets consisted of sentences selected from the TIMIT training database and test database, respectively (TIMIT, 1990). In the training set, eight sentences spoken by 32 speakers were used. In the test set, eight sentences spoken by 16 speakers were used. The signals were low-pass filtered at 4 kHz. The tenth-order LPC analysis is achieved for each 20 ms segment by using the autocorrelation method with Hamming-windowed speech signals. A total of 38038 LSF vectors and 19998 LSF vectors are used for training and testing purposes, respectively. The speech frame is labeled as either unvoiced or voiced following the procedure presented in (ISO, 1998). The class statistics of the training set possess 55% unvoiced class and 45% voiced class whereas the class statistics of the test set consist of 53% unvoiced class and 47% voiced class.

Several structured LSF quantizers with different bit rates, codebook structures and numbers of pre-selected codewords for the M-L tree search are adopted in simulations. For example, the notation ‘CQ-B24-(6)-(9,8)-M2’ means that the classified VQ is classified at the bit rate 24 bits/frame, six bits for the first stage without splitting and two-SVQ for the second stage where nine bits and eight bits are assigned for the first and second subvectors, respectively. Finally, the notation ‘M2’ denotes that \(M\), the number of pre-selected codewords for the M-L tree search is set as 2. If the SVQ scheme is adopted, the 10-dimensional LSF vector is divided into two subvectors of five dimensions or is divided into three subvectors with \((3, 3, 4)\) splitting.

First, the performance of LSF VQ with and without class classification is compared. The CQ (i.e. the proposed CLPQ scheme) and UQ denote whether the LSF vector is quantized with or without classification, respectively. Tables 1 and 2 show the performance of the LSF quantizer for the voiced frames and unvoiced ones, respectively. In the multi-model speech coders, the extra bit required for the class of each speech frame has been shared with other parts of the speech coding. Based on knowledge of information theory, if the voiced frame and unvoiced frames are equally likely, the CQ can save one bit. From the results illustrated in Table 1 (bit rate 22-24 and \(M = 4\)), we can find that the CQ successfully reduces the bit rate by about 1 bit/frame. It is noted that, for the training set, the CLPQ scheme achieves the transparent coding of LPC spectra at a bit rate about 21 bits/frame and 10 bits/frame for the
voiced and unvoiced frames, respectively. Note however, that the bit rate of transparent coding for the test set is not as low as the training set. We believe that the performance of LSF VQ for the test set can be further improved if a larger training set is adopted or if the number of pre-selected codewords, $M$, is increased. The VSLQ scheme achieves transparent coding of LPC parameters at the bit rate 25 bits/frame. The proposed CLPQ scheme achieves transparent coding at a lower bit rate than the VSLQ scheme.

Memory requirement and computational complexity depend on the codebook structure. The computational complexity is also related to $M$. If the first stage is the non-split VQ with bit assignment $b_{1,1}$ and the second stage is the SVQ where the residual vector is split into $n$ subvectors with dimensions $(d_{2,1}, d_{2,2}, \ldots, d_{2,n})$ and each subvector is quantized with bit assignment $(b_{2,1}, b_{2,2}, \ldots, b_{2,n})$, the memory requirement $MR$ is

$$MR = L_{1,1} \times N_p + 2 \times \sum_{i=1}^{n} L_{2,i} \times d_{2,i}, \quad (10)$$

where $L_{1,1} = 2^{b_{1,1}}$ and $\sum_{i=1}^{n} d_{2,i} = N_p$. The computational complexity is

$$C = L_{1,1} \times 2N_p + 4M \times \sum_{i=1}^{n} L_{2,i} \times d_{2,i}, \quad (11)$$

where $C$ denotes the number of multiply-add operations per frame.

Figure 2 depicts the number of pre-selected codewords, $M$, in the M-L tree search scheme versus the average $SD$ of voiced frames for different CQ structures. For each configuration, the different number of $M$ ($= 1, 2, 4, 8, 16, \ldots$) is analyzed. As $M$ is increased, the spectral distortion is reduced accordingly. Table 3 shows the memory requirement and computational complexity with different structures. In (Paliwal and Atal, 1993), a universal quantizer with one stage two-split VQ achieves transparent coding at the bit rate 24 bits/frame and requires about 40K of memory and about 80000 multiply-add operations per frame. The CLPQ scheme, with a configuration denoted as ‘B23-(6)-(8,8)-M4’, has a smaller memory requirement, less computational complexity and fewer bits. The computational complexity of the LSF quantizer, at the bit rate 21 bits/frame, presented in (Hiwasaki et al., 2004) is about 3 million operations. The CLPQ scheme with a ‘B21-(10)-(10)-M32’ configuration requires less computational complexity.

### Table 1 Performance comparison between UQ and CQ schemes (voiced frame)

<table>
<thead>
<tr>
<th>LSF VQ Design</th>
<th>UQ SD, dB</th>
<th>Outlier, % &gt; 4dB, %</th>
<th>CQ SD, dB</th>
<th>Outlier, % &gt; 4dB, %</th>
<th>UQ SD, dB</th>
<th>Outlier, % &gt; 4dB, %</th>
<th>CQ SD, dB</th>
<th>Outlier, % &gt; 4dB, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>B24-(6)-(9, 8)-M2</td>
<td>1.08</td>
<td>0.05</td>
<td>1.03</td>
<td>0.48</td>
<td>0</td>
<td>1.05</td>
<td>0.81</td>
<td>0</td>
</tr>
<tr>
<td>B24-(5)-(6, 6, 8)-M8</td>
<td>1.04</td>
<td>0.31</td>
<td>0.97</td>
<td>0.65</td>
<td>0</td>
<td>1.02</td>
<td>0.95</td>
<td>0</td>
</tr>
<tr>
<td>B24-(6)-(9, 8)-M4</td>
<td>1.02</td>
<td>-</td>
<td>0.96</td>
<td>-</td>
<td>-</td>
<td>0.99</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B23-(6)-(8, 8)-M4</td>
<td>1.07</td>
<td>1.15</td>
<td>1.02</td>
<td>0.55</td>
<td>0</td>
<td>1.05</td>
<td>0.75</td>
<td>0</td>
</tr>
<tr>
<td>B22-(10)-(11)-M4</td>
<td>1.15</td>
<td>-</td>
<td>1.10</td>
<td>-</td>
<td>-</td>
<td>1.08</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B22-(10)-(11)-M8</td>
<td>1.10</td>
<td>1.29</td>
<td>1.05</td>
<td>0.86</td>
<td>0</td>
<td>1.04</td>
<td>0.75</td>
<td>0</td>
</tr>
<tr>
<td>B21-(10)-(10)-M32</td>
<td>1.09</td>
<td>1.81</td>
<td>1.04</td>
<td>1.11</td>
<td>0</td>
<td>1.04</td>
<td>1.05</td>
<td>0</td>
</tr>
<tr>
<td>B21-(10)-(5, 5)-M64</td>
<td>1.16</td>
<td>3.21</td>
<td>1.06</td>
<td>1.61</td>
<td>0</td>
<td>1.09</td>
<td>1.50</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 2 Performance comparison between UQ and CQ schemes (unvoiced frame)

<table>
<thead>
<tr>
<th>LSF VQ Design</th>
<th>UQ SD, dB</th>
<th>&gt; 4dB, %</th>
<th>CQ SD, dB</th>
<th>&gt; 4dB, %</th>
<th>UQ SD, dB</th>
<th>&gt; 4dB, %</th>
<th>CQ SD, dB</th>
<th>&gt; 4dB, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>B12-(5)-(6)-M8</td>
<td>2.05</td>
<td>0.50</td>
<td>2.00</td>
<td>0.53</td>
<td>2.03</td>
<td>0.56</td>
<td>1.97</td>
<td>0.40</td>
</tr>
<tr>
<td>B12-(5)-(6)-M16</td>
<td>2.01</td>
<td>0.43</td>
<td>1.96</td>
<td>0.47</td>
<td>1.99</td>
<td>0.49</td>
<td>1.94</td>
<td>0.40</td>
</tr>
<tr>
<td>B11-(5)-(5)-M32</td>
<td>2.12</td>
<td>0.86</td>
<td>2.06</td>
<td>0.83</td>
<td>2.07</td>
<td>0.81</td>
<td>2.04</td>
<td>0.82</td>
</tr>
<tr>
<td>B11-(5)-(5)-M16</td>
<td>2.15</td>
<td>0.97</td>
<td>2.12</td>
<td>0.95</td>
<td>2.12</td>
<td>0.95</td>
<td>2.10</td>
<td>0.98</td>
</tr>
<tr>
<td>B10-(10)-M1</td>
<td>2.25</td>
<td>1.06</td>
<td>2.22</td>
<td>1.12</td>
<td>2.05</td>
<td>0.70</td>
<td>1.96</td>
<td>0.77</td>
</tr>
</tbody>
</table>

### V. CONCLUSIONS

In this paper, a CLPQ scheme was designed to quantize the LSF vector with phonetic classification. The experimental results show that unvoiced spectra can be quantized with 10 bits/frame whereas voiced spectra can be quantized using bit rates as low as 21 bits/frame. Hence the CLPQ scheme quantizes the
LPC parameter with a variable bit rate. Benefiting from the M-L tree search scheme, the proposed scheme also provides computational scalability. Many classified LSF quantizers with different codebook structures and bit rates were evaluated. The proposed scheme was suitable for the modern hybrid speech coder, which encodes the particular class speech with a different coding strategy at a low bit rate.

ACKNOWLEDGMENTS

This research was supported by the National Science Council under Contract #NSC-95-2221-E-244-008, Taiwan, R.O.C.

REFERENCES


TIMIT, 1990, DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (CD-ROM), National Institute of Standards and Technology, NIST Speech Disc 1-1.1, USA.