ROBUST SPEECH MODE BASED LSF VECTOR QUANTIZATION FOR LOW BIT RATE CODERS

S. Nandkumar, K. Swaminathan, U. Bhaskar

Hughes Network Systems
11717 Exploration Lane
Germantown, Maryland 20876, USA

ABSTRACT

Robust vector quantization of LSF parameters at a low bit rate is essential for voice coders operating below 5 Kbps. A novel aspect of the proposed technique is the use of decorrelated residual LSF vectors from speech mode based backward prediction along with a multi-stage VQ design. Rates as low as 12 bits per 20 ms speech frame for the stationary voiced speech mode and 22 bits/frame for unvoiced and non-stationary voiced frames are shown to result in efficient quantization. In our classification scheme, spectrally stationary voiced frames constitute around 30% of active speech frames resulting in a minimum average bit rate of 19 bits/frame. Objective VQ performance is compared with cellular standard coders such as IS-641 and IS-127. The proposed VQ has been integrated into a speech mode based 4.8 Kbps coder resulting in subjective performance close to that of the 7.4 Kbps IS-641 coder.

1. INTRODUCTION

Speech coders with bit rates ranging from 3000 to 5000 bps are desirable for effective bandwidth utilization in wireless communications applications. Efficient vector quantization of spectral parameters of speech is essential for high quality low bit rate linear prediction based speech coders. Performance of the vector quantizer should be robust across speaker and channel variabilities, and fairly resistant to transmission bit errors. Efficient VQ design schemes for LSF parameters of speech have been proposed in the past [1, 2]. The Split Vector Quantization (SVQ) technique proposed in [1] is shown to result in efficient quantization at 24-26 bits per 20 ms frame. The IS-641 coder (a TDMA cellular standard) uses a 3-way 26 bit SVQ with first order backward prediction to effectively encode LSF parameters. Another cellular standard coder, the variable rate IS-127 uses a 28 bit 4-way SVQ for the full-rate (8000 bps) option, and a 22 bit 3-way SVQ for the half-rate (4000 bps) option. The Multi-Stage Vector Quantization (MSVQ) scheme presented in [2] is also shown to result in efficient quantization performance at 22-24 bits per 20 ms frame. Furthermore, the multi-stage structure has more flexibility in terms of search complexity, codebook storage, and channel error protection. In the proposed quantization scheme, a multi-stage VQ structure is proposed based on the speech mode (spectrally stationary and voiced, spectrally non-stationary or unvoiced). A front-end mode classification scheme identifies the speech frame as belonging to one of three modes (voiced stationary, voiced non-stationary, and unvoiced/silence/noise). The parameters that are quantized are residual Line Spectral Frequency (LSF) parameters of speech. The residual LSF vectors are obtained using first order backward prediction involving past speech frames. During the spectrally stationary voiced speech mode, experimentally derived optimal backward prediction coefficients are used to decorrelate the highly time-correlated LSF parameters, and the resulting LSF residual vectors are shown to be efficiently quantized using 12-14 bits per frame. During all other modes where the degree of correlation between the LSF parameters over time is much lesser, a small fixed constant is used for backward prediction and the resulting LSF residual vector is shown to be efficiently quantized using 22-24 bits per frame.

The proposed Multi-Mode Multi-Stage Vector Quantization (MM-MSVQ) scheme is presented in detail as follows. The front-end processing and MM-MSVQ design and search procedures are explained in Section 2. Objective and subjective performance for the MM-MSVQ scheme is presented in Section 3. An example of using such a quantization scheme with a low bit rate speech coder is briefly explained in Section 4, followed by conclusions.

2. MM-MSVQ DESIGN

The MM-MSVQ codebooks are trained using a large set of speech data made up of various talkers, different acoustic backgrounds and channel filters (approximately 1.3 million vectors). The parametric representation of speech in this case is the residual LSF vector which is formed based on the speech mode after first order backward prediction. The mode classification scheme, the mode based derivation of the LSF residual vector, and the codebook design and search techniques are explained in the following sub-sections.

2.1 Mode Classification

In this application, modes assigned to each frame of speech are broadly defined as spectrally stationary voiced mode (mode A), non-stationary voiced mode (mode B), unvoiced/silence/noise mode (mode C). The classification scheme is designed to be consistent across varying levels of speech and background noise. The pattern classification problem is solved using a fuzzy systems approach. The background noise level is updated every frame based on the probability of voice activity in that frame and a set of thresholds are adapted to this noise estimate. The fuzzy system operates on open-loop pitch deviation from past frame, cepstral distance from past frame, energies of four equally divided frequency subbands and assigns a certain weight to the input frame regarding its membership in one of the three modes. A voice activity indicator is also available as an output of the mode classification. This mode classification scheme is primarily used as a front-end for a low bit rate mode based speech coder.

The spectrally stationary Mode A frames will be used to train a 2-stage MSVQ and the Mode B and Mode C frames will be used to train a 4-stage MSVQ.

2.2 Mode Based Residual LSF Vector

The LSF parameter vector is obtained by transforming a 10th order LPC parameter vector. Next, the long-term average LSF vector (obtained by averaging the LSF vectors in the training set) $\mathbf{l}_{DC}$ is subtracted from the LSF vector belonging to the *i*th frame $\mathbf{l}_i$ to obtain a differential LSF vector $\mathbf{d}_i$ given by

$$\mathbf{d}_i = \mathbf{l}_i - \mathbf{l}_{DC}.$$

During the spectrally stationary mode A frames, adjacent LSF vectors are highly correlated. This correlation can be removed using first-order backward prediction, where the correlation coefficients are represented by a diagonal matrix $\hat{\mathbf{A}}$. The correlation coefficients are estimated in a minimum mean square error sense from the training set of differential LSF vectors classified as mode A, and the diagonal elements of $\hat{\mathbf{A}}$ are given by

$$\hat{\mathbf{A}}[j][j] = \frac{\sum_{i=1}^{N} \mathbf{d}_i[j] \cdot \mathbf{d}_{i-1}^T[j]}{\sum_{i=1}^{N} \mathbf{d}_{i-1}^2[j]},$$

where *N* is the number of frames in the training set. Next, the LSF residual vector for the *i*th frame $\mathbf{e}_i$ in the case of spectrally stationary mode A frames, is obtained as

$$\mathbf{e}_i = \mathbf{d}_i - \hat{\mathbf{A}} \cdot \mathbf{d}_{i-1}.$$

During non-stationary voiced (mode B) and unvoiced (mode C) frames, adjacent frame LSF vectors are not well correlated. Hence an experimentally determined scalar quantity α with a low value of *0.375*, is used as the correlation coefficient for backward prediction of the LSF residual vector. In this case, the LSF residual vector is given by

$$\mathbf{e}_i = \mathbf{d}_i - \alpha \cdot \mathbf{d}_{i-1}.$$

2.3 Weighted Distortion Measure

A weighted Euclidean distance measure similar to the one proposed in [1] is used for training the multi-stage codebooks and during the search for the best codevector during quantization. The weighted distortion measure $d(\mathbf{e}, \hat{\mathbf{e}})$ between the input LSF residual vector $\mathbf{e}$ and the quantized LSF residual vector $\hat{\mathbf{e}}$ is given by

$$d(\mathbf{e}, \hat{\mathbf{e}}) = \sum_{j=1}^{p} w_j (e_j - \hat{e}_j)^2$$

where *p* (*p = 10* in our case) is the number of elements in the LSF residual vector, and w_j is the weight assigned to the *j*th Line Spectral Frequency. The weight w_j is given by evaluating the LPC power spectrum at the *j*th Line Spectral Frequency l_j, and raising it to the power of 0.3 as given in [1].

2.4 Multi-Stage Codebook Design

The iterative sequential design technique used in this study to train the multi-stage VQ's is similar to the one described in [2]. The iterative sequential design technique consists of two steps. The first step involves designing an initial set of multi-stage codebooks in a sequential manner. Here the sequential design implies that the codebook at each stage is designed using a training set consisting of quantization error vectors from the previous stage. It is evident that the codebook at the first stage uses the training set of LSF residual vectors. The codebooks at each stage are trained using the well known generalized Lloyd algorithm with the weighted distortion as defined in the previous section. In this first step of multi-stage VQ design, it is assumed at each stage that all the following stages consist of null vectors.

The second step of the iterative sequential design involves iterative re-optimization of each stage in order to minimize the weighted distortion over all stages. Since an initial set of multi-stage codebooks are known, each stage is optimized given the other stages. In other words, the training set for each stage is the quantization error between the input LSF residual vector and a reconstruction vector which consists of minimum distortion codebook vectors from all stages except the one being re-optimized. This process is continued iteratively until a pre-defined convergence criterion is met. For spectrally stationary mode A frames, a two-stage 12 bit VQ (64 vectors per stage) and a two-stage 14 bit VQ (128 vectors per stage) are designed using a subset of the training data that are classified as mode A (about 340,000 vectors). For all other modes, a four-stage 22 bit VQ (64 vectors per stage for the first two stages, and 32 vectors per stage for the last two stages) and a four-stage 24 bit VQ (64 vectors per stage) are designed using the entire training data set (about 1.3 million vectors).

2.5 Codebook Search and Quantization

During training and encoding, the multi-stage codebooks are searched using an M-L tree search procedure as described in [2]. In the M-L search procedure, considering the codebook at the first stage, M (M = 8 is experimentally found to be sufficient) codebook vectors which achieve the lowest weighted distortion are first selected. Next, M quantization error (from first stage) vectors are computed and the second codebook is searched using the M error vectors and M paths that achieve the overall lowest distortion are selected. This procedure is continued for all stages of the codebook. After M paths are obtained for all the stages, the best out of M paths is chosen by minimizing the weighted distortion measure between the input LSF residual vector and the overall quantized vector. The overall quantized vector for each given path is the sum of codevectors from all stages of the codebook. The indices of the chosen codevector from each stage is transmitted to the speech decoder, and the quantized LSF vector is reconstructed as follows,

$$\mathbf{l}_i^q = \mathbf{e}_i^q + \mathbf{l}_{DC} + K\left(\mathbf{l}_{i-1}^q - \mathbf{l}_{DC}\right),$$

where $\mathbf{l}_i^q$ is the quantized LSF vector at the *i*th frame, $\mathbf{l}_{i-1}^q$ is the quantized LSF vector at the *i-1*th frame, $K = \hat{\mathbf{A}}$, a diagonal matrix during mode A, and $K = \alpha$ a scalar during modes B and C. The MM-MSVQ quantization and reconstruction scheme is illustrated in Figure 1.

3. PERFORMANCE ANALYSIS

In this section, results are presented for the proposed MM-MSVQ scheme. The test data used for these results are separate from the large set of speech data that was used to train the MM-MSVQ codebooks and includes speech utterances at different levels and noise cases. The test speech data (greater than 100,000 frames) is passed through the front-end mode classification scheme and quantized LSF vectors are reconstructed using the MM-MSVQ codebooks. The quantized and original LSF vectors are compared using averages and outlier percentages of the well known log spectral distortion metric [1]. It is known that for efficient quantization, an average log spectral distortion of 1 dB across all test vectors is very important [1]. In Table 1, log spectral distortion statistics are presented for the 12/22 bit MM-MSVQ codebooks (average bit rate = 19) and compared to the performance of a 22 bit Split VQ codebook which has been used in the half-rate operation of the IS-127 coder (LSD refers to Log Spectral Distortion over the entire frequency band of 0-4 Khz for 8 Khz sampled speech, and LSD1 refers to the frequency band of 0-3 Khz which contain more of the high formant energies). It can be clearly seen that for the 22 bit split VQ, the average log spectral distortion is greater than the 1 dB criterion by 0.56, whereas for the 12/22 bit MM-MSVQ codebooks the average log spectral distortion is maintained at 1.11 dB. Moreover, outliers in the range of 2-4 dB are at 9.99% for the 22 bit split VQ, whereas for the 12/22 bit MM-MSVQ the same outliers make up around 3.18% of all test vectors.

Performance Measure	12/22 bit MM-MSVQ	22 bit Split VQ (IS-127)
Average LSD	1.11	1.56
% frames > 2dB	3.18	9.99
% frames > 4dB	0.02	0.02
Average LSD1	1.10	1.60
% frames > 2dB	2.97	13.99
% frames > 4dB	0.035	0.05

Table 1. Performance of 12/22 bit MM-MSVQ versus 22 bit Split VQ from IS-127 half-rate operation.

Results are also presented for the higher rate 12/24 bit MM-MSVQ (average bit rate = 20) and the 14/24 bit MM-MSVQ (average bit rate = 21) versus the 26 bit Split VQ of the IS-641 standard and the 28 bit Split VQ of the IS-127 full-rate operation in Table 2. It is seen that performance of the 12/24 bit MM-MSVQ is comparable to the higher rate 26 bit SVQ, and the 14/24 MM-MSVQ is comparable to the higher rate 28 bit SVQ.

Quantization Scheme	LSD	% > 2 dB	% > 4 dB
12/24 bit MM-MSVQ	1.03	2.22	0.03
26 bit Split VQ (IS-641)	1.09	2.23	0.01
14/24 bit MM-MSVQ	0.98	1.65	0.02
28 bit Split VQ (IS-127 full rate)	1.14	1.55	0.01

Table 2. Performance of 12/24 bit and 14/24 bit MM-MSVQ versus 26 bit and 28 bit Split VQ from IS-641 and IS-127 full-rate operation respectively.

The structure of the MSVQ lends itself to more efficient bit error protection based on the fact that the initial stages are expected to be more sensitive. For example, an FEC scheme can focus on correcting the more sensitive bits and leave the less sensitive bits unprotected. In order to support this claim and as an aid to a bit selective FEC scheme, results from a bit sensitivity analysis using the 2-stage 12 bit MM-MSVQ (mode A) and the 4-stage 22 bit MM-MSVQ (modes B and C) are presented in Table 3 and Table 4 respectively. Bit errors are introduced one bit at a time (error rate = 10 %) and are presented in the tables starting from MSB error to LSB error. The average LSD and outlier performance indicate that errors affect bits from the later stages less than bits from the earlier stages.

Bit Errors	Av. LSD	% > 2dB	% > 4dB
0 Errors	1.27	4.4	0.0
I - B1	1.62	20.9	2.66
I - B2	1.67	21.3	3.9
I - B3	1.60	19.8	2.15
I - B4	1.57	19.4	1.3
I - B5	1.48	16.1	0.2
I - B6	1.42	11.7	0.01
II - B1	1.47	15.2	0.08
II - B2	1.46	14.5	0.07
II - B3	1.47	15.5	0.09
II - B4	1.46	14.4	0.05
II - B5	1.44	13.5	0.05
II - B6	1.43	12.1	0.07

Table 3. Bit sensitivity analysis for a 2-stage 12 bit MM-MSVQ (mode A).

4. LOW BIT RATE CODER APPLICATION

The proposed MM-MSVQ has been integrated into a 4.8 Kbps multi-mode CELP based coder. Bit allocations for the various encoder parameters are chosen in a mode specific manner in order to achieve optimum performance for each mode. The low bit rate MM-MSVQ enables optimum allocation of bits to excitation and pitch parameters especially in Mode A. The reader is referred to [3] for further details. The 4.8 Kbps coder is seen to perform consistently across speech levels and channel conditions. Subjective MOS scores confirm that performance under clear

channel conditions are close to that for the IS-641 7.4 Kbps coder [3].

Bit Errors	Av. LSD	% > 2dB	% > 4dB
0 Errors	1.16	3.6	.03
I - B1	1.91	19.9	9.4
I - B2	1.93	20.5	10.0
I - B3	1.69	17.9	6.8
I - B4	1.52	15.6	4.4
I - B5	1.53	16.0	4.85
I - B6	1.41	13.9	1.6
II - B1	1.51	15.6	4.7
II - B2	1.47	14.9	3.5
II - B3	1.48	15.1	3.8
II - B4	1.44	14.3	2.4
II - B5	1.47	14.9	3.8
II - B6	1.38	13.3	1.06
III - B1	1.30	10.7	0.12
III - B2	1.29	10.3	0.08
III - B3	1.31	11.2	0.12
III - B4	1.30	10.7	0.10
III - B5	1.29	10.4	0.09
IV - B1	1.25	7.27	0.05
IV - B2	1.25	6.96	0.06
IV - B3	1.25	6.9	0.05
IV - B4	1.24	6.75	0.04
IV - B5	1.22	5.6	0.04

Table 4. Bit sensitivity analysis for a 4-stage 22 bit MM-MSVQ (modes B and C).

4. CONCLUSIONS

A novel technique to perform efficient Multi-Stage Vector Quantization of LSF parameters of speech at a low bit rate is proposed. The proposed technique (MM-MSVQ) involves speech mode based MSVQ design using residual LSF vectors obtained from first-order backward prediction of LSF vectors. It is shown that efficient quantization performance can be obtained by designing a 12 bit two-stage codebook for spectrally stationary Mode A vectors and a 22 bit four-stage codebook for unvoiced and spectrally non-stationary voiced frames. The performance of the proposed 12/22 bit MM-MSVQ is compared to a 22 bit Split VQ which is used in the IS-127 standard speech coder and is shown to be superior in terms of average and outlier percentage of the log spectral distortion measure. The proposed MM-MSVQ scheme is also shown to be robust to bit errors and conducive to a bit-selective FEC scheme. Performance of a 12/24 MM-MSVQ and a 14/24 bit MM-MSVQ is shown to be comparable to higher rate VQ techniques such as a 26 bit Split VQ (IS-641 standard) and a 28 bit Split VQ (IS-127 standard). Finally, the proposed MM-MSVQ has been shown to be an integral part of a low bit rate multi-mode coder. Subjective MOS tests have confirmed that the quality of the proposed 4.8 Kbps coder is close to that of the full-rate TDMA standard coder.

5. REFERENCES

[1] K. Paliwal and B. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, No. 1, January 1993.

[2] P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, V. Cuperman, "Efficient Search and Design Procedures for Robust Multi-Stage VQ of LPC Parameters for 4 kb/s Speech Coding," *IEEE Transactions on Speech and Audio Processing*, Vol. 1, No. 4, October 1993.

[3] K. Swaminathan, S. Nandkumar, U. Bhaskar, et al., "A Robust Low Rate Voice Codec For Wireless Communications," *IEEE Workshop on Speech Coding for Telecommunications Proceedings*, September 7-10, 1997.

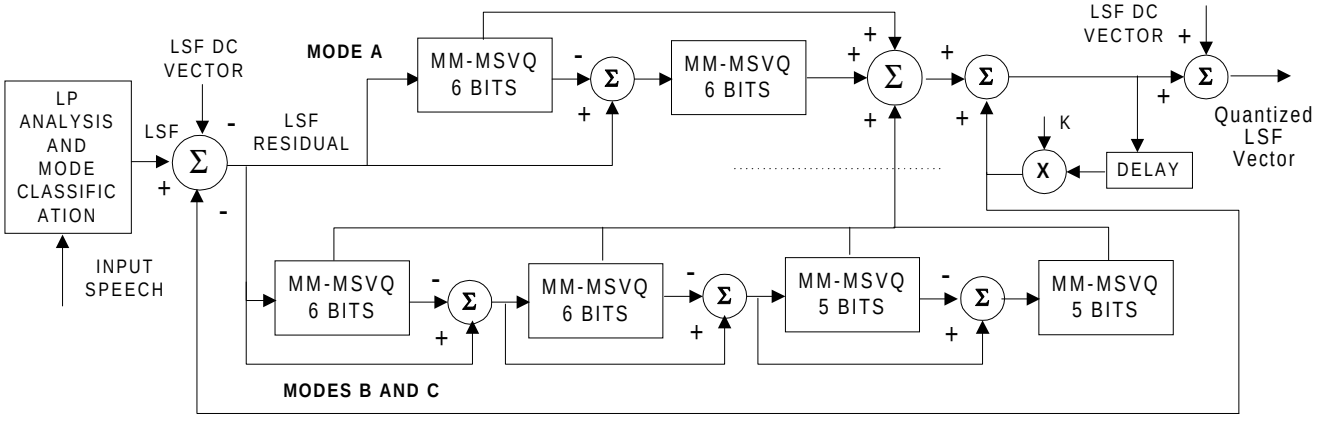

Figure 1. Multi-Mode Multi-Stage Vector Quantization and Reconstruction Scheme.